Consumer Measurement Systems and Child and Adolescent Mental Health

Leonard Bickman
Barry Nurcombe
Clare Townsend
Madge Belle
James Schut
Marc Karver

national
mental
health
strategy

Consumer Measurement Systems for Child and Adolescent Mental Health

LEONARD BICKMAN

BARRY NURCOMBE

CLARE TOWNSEND

MADGE BELLE

JAMES SCHUT

MARC KARVER

University of Queensland

St Lucia, QLD, Australia

And

Vanderbilt University

Nashville TN, USA

ISBN 0 642 36773 6

Additional copies of the report are available from the Mental Health Branch, Commonwealth Department of Health and Aged Care, telephone 1800 066 247 or facsimile 1800 634 400. The report may also be down-loaded from the Mental Health Branch website at:

http://www.health.gov.au/hsdd/mentalhe

Copies of other publications produced under the National Mental Health Strategy are also available at this site.

Authors

Professor Leonard Bickman, M.A., Ph.D., Professor of Psychology, Psychiatry and Public Policy, Vanderbilt University, Centre for Mental Health Policy, Nashville, TN 37212; Bickman@home.com

Professor Barry Nurcombe, M.D., FRACP, FRANZCP, Department of Psychiatry, Director, Child and Adolescent Psychiatry, University of Queensland; bnurcombe@psychiatry.uq.edu.au

Clare Townsend B.A., B.S.W, Research Associate, Department of Psychiatry, University of Queensland; clare@psychiatry.uq.edu.au

Madge Belle B.A., Research Assistant, Department of Psychiatry, University of Queensland; madge@psychiatry.uq.edu.au

L. James A. Schut, M.S., Centre for Mental Health Policy, Nashville, TN 37212

Marc S. Karver, M.S., Centre for Mental Health Policy, Nashville, TN 37212

Research Assistant

Lisa Manning B.A. (Hons)

Acknowledgements

Denisse Best, (Manager), Consumers and Staff of Child and Youth Mental Health Services, the Royal Children's Hospital and District Health Service

Carers Advisory Group, Child and Youth Mental Health Service, the Royal Children's Hospital and District Health Service

John Pearson, (Team Leader) Consumers and Staff, Child and Youth Mental Health Service, Redcliffe-Caboolture District Health Service

Lucia Chang-Tave

Corrine Bickman M.S.

Child and Adolescent Outcome Measurement Project Advisory Group

Bruce Houghton, Contract Manager, Medical Benefits of Australia

Gwenneth Roberts Ph.D.

Commonwealth Liaison

Dermot Casey

Gabriela Taloni

Ian Thompson

Proposed Reference

Bickman, L., Nurcombe, B., Townsend, C., Belle, M., Schut, J., and Karver, M. (1998). *Consumer Measurement Systems in Child and Adolescent Mental Health.* Canberra, ACT: Department of Health and Family Services.

FOREWORD

Under the National Mental Health Strategy, Commonwealth, State and Territory Governments have committed themselves to reforming mental health services in Australia. A key element of the reform agenda is to improve the quality and effectiveness of mental health services. Consumer outcome measurement offers a practical means of promoting this agenda.

In 1997, then Commonwealth Department of Health and Family Services engaged the University of Queensland to review the use of outcome measures in Australia's child and adolescent mental health services. The Consultancy undertook to survey stakeholders and key leaders in the field of child and adolescent services, summarise existing outcome measures and report on the options for the development and implementation of an outcome measurement system. The project was conducted with advice from the Consumer Outcomes Advisory Group.

The study reflects a drive in health services development to measure consumer outcomes effectively and to see the results used appropriately. This drive is part of a wider movement in the community that is linked with concerns that all services be accountable and focused on quality.

Implementation of consumer outcome measurement carries many potential benefits. Measurement systems may provide the means to measure continuously the quality and effectiveness of services. Clinicians and consumers also benefit from increased participation and the review of treatment options and interventions. Further, national measures allow information to be generated and shared between services and across States and Territories.

A crucial factor in the successful implementation of outcome measurement is that consumers and carers, mental health professionals, managers and policy makers understand and support the development of the concept. It is also crucial that they receive the education and resources required to develop and use outcome measures.

In this report, the consultants report on the views and concerns of all the relevant stakeholders. In general, these groups support outcome measurement in child and adolescent mental health services. However, consultations also suggest that stakeholders hold some concerns. For example, some mental health workers fear that consumer outcome measures may not represent the full range of meaningful outcomes. Consumers and carers also emphasise the need to consider consent and privacy issues. Future work will need to be conducted in partnership with stakeholders to ensure such concerns are addressed.

The study recommends the development of a modular outcome measurement system, containing three modules:

- Baseline Follow-up Module which assesses the child's and family's mental health status at entry to the treatment and at later points of treatment;
- Concurrent Module which obtains information throughout the course of treatment; and
- Background Module which assesses factors in the child's and family's background that moderate the course of treatment.

In progressing the report's recommendations, the Advisory Group believes that attention should be directed to developing the Baseline Follow-up module as the first priority. The Advisory Group believes that work on identifying and trialing a clinician rated outcome measure, to provide an additional perspective to the consumer one, is also an important priority.

I would like to thank the members of the Project Advisory Group for overseeing this important initiative. In particular, I would like to acknowledge the contribution of Professor Helen Herrman, Professor George Lipton, Ms Merinda Epstein, Mr Allen Morris-Yates, Ms Karen Connelly, and Ms Julie Vandort. I would also like to thank the project team for their work, in particular, Professor Barry Nurcombe, Ms Clare Townsend, and Professor Leonard Bickman.

Dermot Casey
Acting Assistant Secretary
Mental Health Branch
September 1999

CONTENTS

Page

III CONSULTING THE STAKEHOLDERS *cont'd*

Page

III CONSULTING THE STAKEHOLDERS *cont'd*

EXECUTIVE SUMMARY

A MEASUREMENT SYSTEM FOR CHILD AND ADOLESCENT MENTAL HEALTH

Rationale and Conceptual Basis

The National Mental Health Plan (1992) identified the need for regular reviews of service outcomes, the development of national outcome standards and quality assurance programs to support the reform of Australia's mental health services. An outcome measurement system is integral to these objectives.

The quality of mental health services should be evaluated not in accordance with the number of people served, but rather in terms of the effectiveness of those services. A recent survey of Australian children and adolescents (Zubrick et al., 1995) reveals that relatively few of those with mental health problems had had contact with specialized mental health services in the previous six months. This underutilisation may be partially explained by the relative paucity of existing services. Mental health services for children and adolescents are seriously under-resourced. The shortage of resources may be partially remedied by more efficiently utilizing available assets. Measurement systems have the potential to assist policy makers in the effective reallocation of existing resources. Armed with reliable data concerning service effectiveness, policy makers and administrators will be able to decide rationally how best to allocate resources for the needed expansion of services for children and adolescents. The effectiveness of services is most reliably assessed by the use of standardized measures. However, to be optimally used measurement systems should be integrated into clinical services and policy decisions.

Outcome measurement should be part of a system in which data are systematically collected, recorded, scored, interpreted, and fed back, in timely fashion and appropriate form, to consumers, clinicians, managers, administrators and policy makers. An iterative system of this type is the basis for continuous quality improvement and for rational decisions concerning resource allocation. A *baseline-followup* measurement system, in which information is collected at entry to services and subsequently, is of most use for program management and policy decisions; a *concurrent measurement system*, in which information is collected concurrently with treatment, could provide progressive information relevant to clinical decision making in the individual case; and *a background measurement system* could provide information about those factors that moderate outcome and affect the extent to which the goals of the treatment can be attained.

An ideal measurement system should be sustainable, feasible, comprehensive, flexible, psychometrically sound, developmentally and culturally sensitive, and able to improve clinical effectiveness. Timely information should be collected at *baseline*, *concurrently* with treatment, and at specified *followup* points. A comprehensive system would gather information from multiple informants in in relation to a number ofseveral domains, for example: *the severity and acuity of the child's symptoms; the child's functional impairment; the child's functional strengths; family functioning; the quality of family life; consumer satisfaction; the goals of treatment; the modality, strategy and tactics of treatment; readiness for change; the quality of the therapeutic alliance* and *adherence to treatment*. However, even if the proposed measurement system meets all design criteria, it is not likely to be successful unless the ethical, resource, and political problems of implementing the system are effectively addressed by the education of stakeholders in the design and implementation of the system, and by the recruitment of consumers, carers, clinicians, managers, administrators, and policy makers to collaborate in the design and development of the system.

Methodolgy

Stakeholder Perspectives

Using focus groups, semistructured telephone interviews and mail surveys, the consultancy undertook an extensive survey of the following groups of stakeholders: adolescent consumers, parents, clinicians, state mental health directors, mental health researchers, senior academics, health insurers, and peak organizations in the field of child and adolescent mental health.

A Critical Review of the Literature

Two literature databases were searched for articles, books, book chapters and other sources concerning outcome and process measurement in child and adolescent mental health. 188 instruments were selected for review by excluding the following: any that did not appear in an abstract after 1989; any that are not appropriate for subjects 5-18 years of age or are too restricted in age; any that require more than 30 minutes or advanced education or training to administer; and any that are designed solely for the diagnosis of specific disorders. Personality inventories and projective tests were excluded. The selected measures were classified according to 19 *background*, *process* and *outcome* domains. Each measure was then evaluated in accordance with 29 criteria, related to its psychometric qualities, cultural sensitivity, developmental sensitivity, feasibility and cost. Next, two raters evaluated each measure in accordance with exacting psychometric and feasibility criteria. Each of the "best measures" that escaped the final cull was then critically reviewed with regard to its suitability as a multidimensional measure or as part of a composite measure of treatment outcome.

Results

Stakeholder Perspectives on Outcome Measurement

Although few stakeholders conceived of outcome measurement as a system, the majority endorsed the importance of outcome measurement as a means of enhancing service effectiveness and as the basis for rational resource allocation. Stakeholders were not unanimous concerning the need for a universal or "endorsed" battery of measures: many expressed concern that policy makers might adopt an untested standard system, the erroneous or simplistic information from which would be seriously misleading.

The ideal measurement system was generally described in these terms: congruent with established treatment planning and review; comprehensive; involving multiple informants; brief; "user-friendly", simple to score and interpret; and clinically relevant. The following measurement domains were rated as most important, in order: *family functioning; the quality of the parent-child relationship; the client's level of social functioning; disorder-specific symptomatology; burden of care; the client's physical health and medication; adherence with treatment; parental physical and mental health; global symptom severity;* and *the client's functional strengths*. Respondents were divided concerning the appropriate frequency and timing of data collection. A significant minority of respondents were opposed to concurrent measurement.

Respondents predicted that outcome measurement would be accepted by most stakeholders provided the system was inexpensive, non-burdensome and useful; however, some clinician groups were less than enthusiastic. The chief barrier to implementation was thought to be the resistance of clinicians to intrusion on their clinical autonomy. Respondents suggested that barriers to implementation could be overcome provided the following applied: all stakeholders, particularly clinicians, should contribute to the design, testing, and implementation of the system; all should understand how the data will be used; the data yielded should be accessible, timely, and relevant to clinical practice; and adequate resources should be provided to develop and sustain the measurement system.

Respondents identified the following risks, ethical issues and problems related to outcome measurement: violation of confidentiality; burdensomeness; negative impact on the therapeutic alliance; simplistic, unsophisticated measurement leading to erroneous clinical or policy decisions; threats to funding and resource allocation; and medico-legal problems.

Despite the generally favourable view of outcome measurement (qualified by caution about potential risks and the need for stakeholder involvement at all phases of design, installation, testing and adoption), there were a significant dissenting minority. Their views can be characterized in the following terms. Further discussion is needed concerning the principles, risks and benefits of outcome measurement; otherwise, an inadequate system might be unwisely adopted. Contemporary outcome measurement is primitive and fails to capture the complexity, subtlety, diversity and idiosyncrasy of psychological problems and clinical work. It would take an exceptional effort to establish a measurement system, and there is no guarantee that the mental health of young people would be improved thereby. At most, existing measures could be tentatively used in order to educate clinicians and promote constructive debate.

Review of Literature

Our critical review of the literature identified 31 “best” outcome measurement instruments in the following domains: *Coping* (1); *Family Functioning* (3); *Family Resources* (1); *Functional Competence* (2); *Functional Impairment* (2); *General Symptomatology* (3); *Goals* (1); *Maltreatment* (2); *Quality of Life* (2); *Satisfaction* (1); *Self Esteem* (3); *Social Support* (2); *Stressful Events* (2); *Therapeutic Process* (3); and *Multidimensional Scales* (3).

Measurement Systems Currently Operating in Australia

We were able to locate eleven measurement systems currently operating in this country. Respondents from eight of these systems provided information for our survey.

At the Centre for Developmental Psychiatry, Monash University, Melbourne, baseline-followup measurement systems, tailored for each treatment program, are based on standardized multi-informant checklists and interviews. Outcome data are used to support training, to establish service program profiles, and for funding. The Arndell Children's Unit, Royal North Shore Hospital, Sydney, also uses a multi-informant, multiple-domain, baseline-followup system. Outcome data are fed back to clinicians and managers in the form of computer-generated graphs.

Rivendell Child, Adolescent and Family Services, Sydney uses *satisfaction with services* data to identify cases that have had an unfavourable outcome, for the purpose of quality improvement, *goal attainment* is also assessed. The Child and Adolescent Psychiatry Program, South East Sydney Area Health Service, also uses standard measures of diagnostic, baseline-followup, and background factors for the purpose of case-audited quality improvement, as does the School of Applied Psychology, Griffith University.

Rivendell Child, Adolescent and Family Services, the Alfred Child and Adolescent Mental Health Service, Melbourne, and the Department of Child and Adolescent Psychiatry, Royal North Shore Hospital, Sydney, have implemented baseline-followup systems that incorporate a goal-attainment measure.

Maroondah Child and Adolescent Mental Health Service, Melbourne, is currently piloting the Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA), a new, composite, baseline-followup multiple-domain measure.

It is apparent that a number of Australian services have installed or are piloting measurement systems, predominantly in the baseline-followup mode. The concepts of multiple informants and multiple domains appear to be recognized. However, it is also apparent that relatively few services have installed measurement systems, and that those in use are somewhat patchy in conceptualization and scope.

Conclusions and Recommendations

We propose three modules for development:

- A *Baseline-Followup Module* (BFM), the aggregate data from which can be used to assess service effectiveness.
- A *Background Module* (BM) that provides information to clinicians concerning the moderating variables that potentially impede treatment effectiveness in the individual case.
- A *Concurrent Module* (CM) that provides progressive feedback to clinicians concerning the effectiveness of individual treatment plans.

All these modules could contribute to continuous quality improvement (see appendix 7).

The Baseline-Followup Module

We recommend that the Baseline-Followup Module take less than 30 minutes to complete, involve the clinician minimally, and address the following domains: *functional impairment; symptom severity; symptom acuity; parent-child relationship; quality of life;* and *satisfaction with services*. An existing measure, the Ohio Youth Problems, Functioning and Satisfaction Scales could serve as the core of this module, supplemented by measures of *acuity*, *quality of life*, and *parent-child relationship*. We recommend the use of the Family APGAR measure to assess parent-child relationships and the Students' Life Satisfaction Scales as a measure of quality of life. An acuity measure would need to be developed.

The Background Module

The Background Module should address the following moderating domains: *safety of the environment; stressful events; family resources;* and *maltreatment*. This module would have to be designed and developed de novo.

The Concurrent Module

The Concurrent Module should address the following domains: *the goals of treatment, the modality, strategy, tactics, timing and dosage of treatment;*

therapeutic alliance; motivation to change; adherence to treatment; symptom severity; and *functional impairment.*

We recommend that *goals of treatment* be based on the principles of goal-directed treatment planning. None of the 31 “best instruments” identified from the literature search satisfactorily addresses the *goals of treatment* or describes the *modality*, *strategy*, *tactics*, *timing* and *dosage of treatment*, or *therapeutic alliance*, *readiness for change*, and *adherence to treatment*. We recommend the design, de novo, of instruments for these domains. The Ohio Youth Problems, Functioning and Satisfaction Scales could be adapted to measure change in *symptom severity* and *functional impairment*.

Implementation

We recommend that implementation be incremental and evolutionary. Stakeholders advised us that, unless the measurement system is demonstrably valid and useful, it will not be accepted. In terms of importance and ease of development, we recommend the following priority:

(1) The Baseline-Followup Module

(2) The Concurrent Module

(3) The Background Module

We estimate that initial development and field-testing of the three subsystems will take three years. Mental health services may elect to use any or all of the measurement modules. However, to help ensure adoption, sufficient additional resources should be provided to those agencies that implement these measurement systems.

A number of child and adolescent mental health services are already using the Health of the Nation Outcome Scales-Child and Adolescent (HoNOSCA). The HoNOSCA is described in Appendix 8 of this monograph. We recommend a dual strategy : those services already using the HoNOSCA should continue to do so in parallel with the development of the Baseline-Followup measures recommended in this monograph. Thus, the development of the HoNOSCA will be furthered, and data will be generated for concurrent validation of all measures.

Recent studies (Gowers et al., 1999; Yates, Garralda & Higginson, 1999) which became available after this monograph was prepared provide preliminary support for HoNOSCA's coverage, feasibility, inter-rater reliability, validity as a measure of severity, and sensitivity to change. In order to maintain reliable ratings, it is essential that all raters be trained (and periodically retrained). It has been recommended as a next step that the HoNOSCA be tested against other global severity measures and against child and parent ratings of case severity. More information is required concerning HoNOSCA's developmental and cultural sensitivity, reliability, sensitivity to change, and applicability to inpatient as well as outpatient settings.

Political Issues

Many clinicians are wary of outcome measurement, fearing that it foreshadows cuts in funding, and that, if a simplistic measurement system is prematurely adopted, it will yield misleading information and prompt erroneous policy decisions. Consumers and carers have a vital interest in the availability, accessibility and effectiveness of services and the responsiveness of services to individual needs. However, consumers are unlikely to support outcome measurement unless it is non-burdensome and they think it is useful. Administrators, policy makers, and insurers need to know what services are effective, for whom, and in whose hands. Otherwise, they have no rational basis for resource allocation. They readily endorse the need for a measurement system, provided it is not too expensive.

How can the concerns and needs of the three main stakeholder groups be acknowledged and aligned? First of all, no measurement module should be adopted until it has been thoroughly tested: it should be made clear that we are some years away from a national or “endorsed” system.

All stakeholders need to be aware of each other's concerns and reservations about measurement systems. All need to know that outcome measurement has the potential to improve services and support rational policy decision-making.

Educating the Stakeholders

Innovation requires sustained, committed leadership, a receptive climate among stakeholders, and adequate resources. How can a receptive climate be fostered? The solution is an educational process involving all stakeholders. It should be clear that the design, development, installation and maintenance of a measurement system requires the collaboration of all stakeholders.

Clinicians, consumers and carers are more likely to support outcome measurement if they see the point of it, if it is congruent with the values and aims of the mental health service, and if they perceive it as valuable to them. Concurrent measurement, in particular, has the potential to enhance clinical effectiveness.

The case for innovation should be formally presented to key stakeholders and debated with them. Installation, testing and monitoring should be part of a planned collaborative process, each instrument being modified or refined in accordance with feedback from clinicians, consumers, carers, and administrators. Regular presentations are required concerning the progress of the enterprise, giving all stakeholders the opportunity to contribute. Thus, all stakeholders who contribute will have ownership of the system, and the system itself will engender a spirit of inquiry and self-examination.

CHAPTER I

THE TERMS OF THE CONSULTANCY

On 18th December, 1997, following an Invitation to Tender, the Commonwealth Department of Health and Family Services received a proposal from the authors of this report entitled *Measurement of Consumer Outcomes in Child and Adolescent Mental Health* (Reference: 59/97). Subsequently, the Department awarded a consultancy contract to the University of Queensland. In accordance with the terms of the consultancy, the authors were asked to complete the following tasks:

Task 1 - Review Existing Knowledge

Undertake a literature review to identify instruments potentially useful for outcome measurement.

Task 2 - Survey Stakeholders and Key Leaders

Conduct focus groups, mail surveys, and telephone interviews with carers, consumers, mental health professionals, administrators, policy makers, insurers and key organizations. Ascertain stakeholders' perspectives concerning the purposes and desirability of outcome measurement, the components and characteristics of an ideal measurement system, potential barriers to implementation, and the way in which such barriers might be overcome.

Task 3 - Summarize Outcome Measures

Having located a range of outcome measures from the literature survey, select and describe in detail those instruments that best satisfy the following selection criteria: lack of cultural bias; previous use in Australia; brevity; low cost; lack of a need for training or for advanced qualifications in the administration; suitable reading level; psychometric soundness; up-to-date norms; and developmental sensitivity.

Task 4 - Make Recommendations Concerning the Development and Implementation of a Measurement System

After introducing a conceptual framework concerning an outcome measurement system, describe an ideal system, analyze discrepancies between available instruments and the ideal system, recommend whether the system should be composed of existing instruments or designed de novo, and recommend how the system should be developed and implemented.

Task 5 - Submit a Final Report Incorporating The Results of the Consultancy

This is that report.

CHAPTER II

INTRODUCTION

National Mental Health Guidelines

Adopted by Australian Health Ministers in 1992, the National Mental Health Strategy endorses the promotion of research and outcome evaluation, quality assurance, and national outcome standards. Unquestionably, the development of a measurement system is important to the realization of the National Mental Health Strategy. Indeed, outcome measurement is specified in the following objective of the Strategy:

Objective 30	To institute regular review of outcomes of services provided to persons with serious mental health problems and mental disorders as a central component of mental health service delivery.

Outcome measurement is also embodied or implied in other objectives, as follows:

Objective 31	To develop a national mental health data strategy.
Objective 32	To encourage the development of national outcome standards for mental health services and systems for assessing whether services are meeting these standards.
Objective 33	To ensure all mental health services have quality assurance programs.

The National Standards for Mental Health Services (1997) are described as “outcome oriented with an emphasis on the end result for consumers and carers” (p.1). Mental Health Services are expected to have a strategic plan involving a “service evaluation plan including the measurement of health outcomes for individual consumers” (p.21) and to “routinely monitor health outcomes for individual consumers using a combination of accepted quantitative and qualitative methods” (p.23). Furthermore, prior to each consumer’s exit from the mental health system, staff are expected to “review the outcomes of treatment and support” and to monitor performance in regard to this criterion and others (p.47).

Treatment outcome evaluation is integral to the National Mental Health Strategy (1992) and the National Standards for Mental Health Services (1997). What progress has been made since 1992 with regard to the design and implementation of outcome measurement? In 1994, Andrews, Peters and Teeson published *The Measurement of Consumer Outcomes in Mental Health*, a review of existing outcome measures in adult mental health, with recommendations concerning the measures most suitable for implementation. These measures were subsequently piloted and the results reported in *Measuring Consumer Outcomes in Mental Health* (Stedman, Yellowlees, Mellsop, Clarke & Drake, 1997).

The present report will address conceptual issues associated with process and outcome measurement in Child and Adolescent Mental Health, summarise the opinions of stakeholders concerning the advisability of implementing a measurement system, critically review existing measurement instruments, describe several measurement systems already operating in Australia, and conclude with recommendations about the measurement system most suitable for this country.

The Conceptual Basis of the Current Review

Who is Served by the Child and Adolescent Mental Health System?

In a Western Australian survey of children and adolescents, Zubrick et al., (1995) found that the prevalence of mental health problems was 17.7%. Of the number who were judged to need services only 2% had had contact with specialised mental health services in the six months prior to the survey. Many more had contact with general practitioners, school teachers, and school guidance counsellors. Children with somatic complaints, aggressive behaviour, antisocial behaviour, and social problems were more likely than those with depression, anxiety or social withdrawal to be referred to specialised services. Zubrick et al., suggest that these findings indicate the need for a wiser allocation of resources, with “improved targeting and coordination of mental health treatment and support services currently provided through the health, education, community services and justice sectors”. It is evident from this study that emotionally disturbed children are seriously underserved.

How Are They Served?

Child and Adolescent Mental Health Services are usually provided to patients up to the age of leaving secondary school. Services are ideally delivered in a coordinated fashion within the following possible levels of care:

- Inpatient hospitalization
- Partial hospitalization
- Residential care
- Therapeutic foster care
- Day treatment
- Outpatient services
- In-home services
- Outreach services
- Consultation-liaison services

Each mental health service is usually composed of a multidisciplinary team pooling the skills of different professionals. Mental health services have local managers and are coordinated, usually, by regional administrators and policy makers. Child and adolescent mental health services are usually administratively integrated within regional mental health services that are, in turn, part of regional health services. Child and adolescent mental health services have more or less coordinated links to agencies outside the health system (for example, to the educational and juvenile justice systems) from whom patients are referred or to which consultation services are provided. Public child and adolescent mental health services relate, also, to mental health clinicians in the private sector, and to general practitioners and paediatricians who provide primary care to many children and adolescents, and who refer complex cases to mental health services or share their patient's care with those services.

In Australia, few child and adolescent mental health services provide the full range of care described above. Day treatment, residential care, therapeutic foster care, outreach services and specialised hospital beds are scarce. Given the known need, even outpatient services are seriously under-resourced. Without information on service effectiveness, administrators will have difficulty allocating resources for service expansion or development and staff training.

The sequential clinical processes undertaken by solo mental health clinicians or (in inpatient services) by teams of clinicians are as follows:

1. Referral
2. Intake
3. Diagnostic evaluation
4. Diagnostic formulation
5. Treatment planning
6. Negotiation of the diagnostic formulation and treatment plan with the patient and family.
7. Implementation of treatment
8. Termination.

Aside from clinical interviews, physical examinations, and special investigations (e.g., speech and language evaluation, electroencephalogram), standardised measures (e.g., the Child Behaviour Checklist) may form part of stage 3, *diagnostic evaluation*. However, this report is concerned with stage 7, *the implementation of treatment*, and specifically with the design of a measurement system that tracks the quality of implementation and gauges the effectiveness of services. It is conceivable that some instruments designed for diagnostic use could also prove useful in outcome measurement; however, diagnosis should be distinguished from outcome evaluation and quality improvement.

Why Measure Outcome?

The quality of a mental health service is customarily evaluated in terms of the number of people served and the number of services delivered. In Australia today, there are no bases for administrative decision making other than fiscal restraint, conservatism, lobby group pressure, political opportunism, ideology, or fashion. The aim of outcome measurement, in contrast, is to foster decision-making in terms of service effectiveness, thus to improve the efficiency and effectiveness of services. Theoretically, service effectiveness could be promoted in several ways. First, armed with reliable data, policy makers could make rational decisions concerning the allocation of resources (and track the effectiveness of their decisions). Second, outcome data could help clinicians and local managers decide which treatment is most effective for a particular kind of case, or whether a particular kind of treatment needs modification. Third, provided with timely data about patient progress during service delivery, a clinician would know whether a treatment plan is working and, if appropriate, alter or modify the plan accordingly. Lastly, a measurement system could be used to match patients who have a particular problem with clinicians who are known to be especially effective in treating that kind of problem. More specific uses for each stakeholder group are described below.

This report identifies three different measurement systems for use in child and adolescent mental health services. First, a *Background Module* that assesses factors in the child and family's background that moderate or affect the course of treatment. Second, a *Baseline-Followup Module* that assesses the child and family's mental health status upon entry to services and then at a later point or points. Finally, the *Concurrent Module* which is used to obtain critical information progressively during the course of treatment. A similar classification of data systems is made by Barkham *et al*. (1996).

The purpose of a measurement system is to produce useful information. Two things are required to design a measurement system that yields informative data. Firstly, usefulness should be built into the system. Our emphasis on stakeholder involvement contributions should assure the design of a truly informative system. Secondly, the continuing incremental evaluation and updating of the system is essential to ensure that it continues to serve the needs of key stakeholders. All measurement systems should be continually evaluated. This section discusses the way in which different stakeholders could use the information yielded by a reliable and valid measurement system.

Consumers

A *baseline-followup subsystem* can be used by consumers to review the effectiveness of different mental health services. A *concurrent subsystem* is also potentially useful for consumers : clinician and client can jointly review the progress of treatment.

Clinicians

A comprehensive baseline data collection procedure can provide the clinician with information useful for treatment planning. A *concurrent subsystem* provides systematic information about patient progress. By this means, the clinician can judge the effectiveness of the individual treatment plan and adjust treatment accordingly. Feedback helps the clinician to review treatment goals where progress is not as rapid as predicted.

Managers

Managers of services can use outcome information in several ways. For some purposes they can use data collected at the individual client level. For other purposes data must be aggregated in relation to type of client, type of problem, type of treatment, and clinician. Systematic baseline data can characterise the clientele served by a mental health facility, and determine whether the facility is serving the patients for whom it was designed. By referring to comparable, objective data about the complexity and the difficulty of cases, supervisors can balance clinician's caseloads.

When specific interventions or programs are instituted, treatment effectiveness can be assessed through the use of baseline-followup data. The identification of appropriate treatments, more focused supervision, the targeting of priority consumer groups, clearer decisions about termination of services, the reduction of treatment variability, and improved education and training all contribute to cost efficiency.

The most important benefit of a measurement system is the implementation of continuous quality improvement. Accountability is enhanced when baseline-followup and concurrent subsystems operate.

Policy Makers

Policy makers will aggregate data from several mental health facilities, a procedure requiring computerization and data warehousing. Uniform data collection standards will be required in order to insure that different sites are comparable in the way they collect information. Policy makers make use of data in a way similar to managers, but their scope is broader.

Why Use Standardised Objective Measures?

Clinicians customarily make global, impressionistic assessments of patient progress. Their assessments are not necessarily inaccurate; however, several factors conspire to reduce the reliability of global impressions. In a recent review, Dawes (1994) has concluded that clinical judgment has poor reliability and validity. Several other reviews have indicated flaws in clinical judgment in matters such as the determination of patients' level of functioning, the diagnosis of malingering, personality assessment and the prediction of dangerousness (Douglass, Macfarlane, & Webster, 1996; Kleinmuntz, 1990; Monahan, 1984; Monahan & Steadman, 1994; Rabinowitz, 1993; Rock & Bransford, 1992; Wedding & Faust, 1989). These studies suggest that it is unwise to depend on clinical judgment alone. Clinical judgment should be supplemented by objective measurement.

How Do Measures Designed For Children Differ From Those For Adults?

Research into mental health services for children and adolescents lags behind that in the adult field (Kazdin, 1993a; Knitzer, 1982). One reason for the deficiency is the paucity of reliable, valid and practicable measurement instruments (Bickman, 1996). The measurement instrumentation required for children is more complex than that required for adults: the developmental level of the child, the need for multiple informants, the delivery of services in multiple settings and systems, and the context in which the child lives add dimensions of measurement that are qualitatively different from those for adults. The different developmental stages of childhood affect both the range of normal behaviour and the level and quality of symptoms that indicate psychopathology. Furthermore, children require measurement instruments adapted to their reading ability and level of understanding. Thus, the preparation of measurement instruments appropriate for children requires significantly more than a mere rewording of adult protocols. Outcome and concurrent data should be gathered from multiple informants. Research suggests that, while parents can reliably report children's externalizing behavior (e.g., conduct problems), children are better informants concerning their own internalizing symptoms (e.g., anxiety) and covert actions (e.g., substance use) (Achenbach, McConaughy, & Howell, 1987). Finally, in children, the family is especially important. Information about the family environment is an essential component of any measurement system for children and adolescents. The family can foster development and therapeutic gains (Heflinger & Bickman, 1996). Without information about the family's strengths and needs, clinicians are likely to overlook an important treatment resource.

What Are The Levels Of Measurement?

Beyond consumer outcomes, there is a need for other data that could be used to monitor and improve services. A comprehensive measurement system should include data from the following levels:

1. **Consumer level:** Information about the consumer and his or her progress.
2. **Clinician level:** Information about the provider of mental health services.
3. **Treatment level:** The treatment implemented.
4. **Clinic level:** The extent to which services are used and the cost of those services.
5. **Systems level:** Information concerning the operation of the mental health system as a whole, integrating data collected at subordinate levels and accounting for the relationship of the mental health system to allied systems such as education.

This report will consider a measurement system for consumers from ages 5 to 18 years of age. Future projects need to consider the needs of infants and preschool children. Special attention will also need to be given to Aborigines and Torres Strait Islanders as well as children and their families who come from a non-English speaking background. Other systems of measurement, such as those appropriate to the treatment and systems levels, (e.g. to inpatient hospitalisation) also require study.

The Concept of a System

Outcome measurement involves more than the administration of a measurement instrument. In the ideal situation, data are systematically collected, recorded, scored, and interpreted. Interpretations are conveyed promptly to clinicians, consumers, local managers, regional administrators and central policy makers. Clinical, managerial, administrative and policy decisions are made in the light of interpreted data. The effect of all decisions is tracked. The system is iterative, informing decision-making at the clinical level with the individual patient, at the managerial level with regard to modes of treatment, and at the senior administrative level in regard to services and resource allocation. Iteration is the foundation of continuous quality improvement. See Appendix 6 for a discussion of the application of continuous quality improvement to mental health services.

Since process and outcome information will be used by different stakeholders for different purposes, different types of data are required and different types of measurement will be necessary. A traditional *baseline-followup subsystem*, in which data are collected at the start of services and at least two other times (usually after treatment is completed), is most useful to consumers, managers and policy makers. *Baseline-followup* data provide information about the overall effectiveness of services or programs. If the data are made public, consumers might be able to use the information to select services. However, baseline-follow up data are likely to prove less useful to clinicians in their work with individual consumers because the information is insufficiently fine-grained and unlikely to be provided in a timely fashion. The clinician needs timely data collected concurrently with treatment. *Concurrent information* could include *the consumer's progress towards attaining treatment goals*, *symptomatology*, *adaptive functioning*, and *processes that mediate the success of treatment* (e.g., *the therapeutic alliance* and *adherence to treatment*). In this report we will often refer to these two subsystems: *baseline- followup* and *concurrent*.

Who Are the Stakeholders?

The word “stakeholder” has come to refer to those people, or groups of people, who have a legitimate interest in the running and effectiveness of a mental health system. Based on their prediction of the effect of outcome measurement on the system as a whole, or on their view of the reliability, validity, practicability or utilization of outcome data, different stakeholders are more or less supportive of a measurement system. Among the stakeholders whose opinions have been solicited in this report are the following: adolescent consumers, the parents of psychologically disturbed children and adolescents, clinicians from different disciplines, researchers, service directors who are currently implementing outcome measurement, senior administrators, insurers, and peak bodies. The results of the stakeholder consultations are presented in Chapter III.

Technical Issues

How Often Should Data Be Collected?

It is unquestionably necessary to collect information at the point of entry to services, the *baseline* data-point. However, it is not clear when subsequently to do so. Bickman, Lambert and Schilling (1998) found that it was not possible to predict improvement from baseline data alone. *Concurrent* measurement (i.e., data collected while the child or adolescent is in treatment) is required. Moreover, the points when *followup* information should be collected are unclear. Should data be gathered at fixed intervals or in relation to events like treatment termination? Is the measurement of long-term outcome necessary (Bickman, Lambert & Vides de Andrade, in press). Should the measurement system examine not only short-term change but also longer-term effects? Some children deteriorate four years or longer after starting treatment. Would it be effective to save money by testing only samples of children at followup? The answers to all these questions are not clear. Appendix 6 discusses some of these issues and makes some recommendations.

What Criteria Should Be Followed in the Design of a Measurement System?

Several writers have attempted to define the characteristics of a good measure or measurement system. Green and Newman (1996) list 11 criteria, as follows:

- relevance to target group
- simple, teachable application
- objective referents
- use of multiple respondents
- capacity to describe process
- psychometric soundness
- low cost
- comprehensibility to non-professionals
- uncomplicated interpretation and ease of feedback
- usefulness to clinical services
- compatibility with clinical ideology and practice

The U.S. Center for Mental Health Services has appointed two working groups to study the criteria for the selection of measures. The group focussing on children has advanced the following principles:

- The perspectives of child and family should be included.
- The system should be compatible with community values.
- The system should be easy to use.
- The system should address the goals of the program or intervention.
- Measures should be reliable and valid.
- Indicators of physical health, environmental safety, educational achievement, and legal status should be included.
- Data should be collected at regular intervals, and include those who drop out of treatment.
- Data should be collected concerning treatment modality, fidelity of treatment, and quality of services.

A second group at the U.S. Center for Mental Health Services is working on the technical standards for measurement. When possible, the standards propounded by this group have been incorporated in this report as follows:

- Feasibility
- Comprehensiveness
- Flexibility
- Potential to improve clinical effectiveness
- Psychometric soundness
- Developmental sensitivity
- Cultural sensitivity.

Feasibility: Feasibility is the most exacting criterion. We have adopted the definition used by Stedman et al. (1997) in their report on adult measures, after Andrews, Peters and Teeson (1994): feasibility is the extent to which a measure is applicable, acceptable, and practicable. Instruments should be easy to administer in clinical settings and take little effort and time for respondents to complete. Little or no training should be required for administration, and the data yielded should be easy to analyse and interpret. Timely information is required at specified points in the clinical process, for example:

Baseline	(initial severity, complexity and likelihood of change, treatment planning)
Concurrent	(progress, feedback to clinician and managers)
Discharge	(short-term outcome, discharge planning)
Followup	(long-term outcome, discharge planning)

Our review of extant instruments in Chapter IV will evaluate the feasibility of each instrument. It is evident from our analysis that almost all these instruments have been designed for research and not for application to outcome measurement in the clinical setting.

Comprehensiveness: A measurement system should address the outcome domains that stakeholders consider important. The assessment of only a single domain (e.g., symptoms) will omit changes that occur in other domains. On the other hand, a comprehensive system will enable the clinician to attend to issues of relevance to the particular case and to assess, for example, whether symptom amelioration is accompanied by improved adaptive functioning. A comprehensive outcome measurement system could encompass the following domains:

- The intensity/frequency of the child’s symptoms (severity)
- The dangerousness of the child’s symptoms (acuity)
- Functional impairment
- Functional strengths
- Family functioning/environmental stability
- Caregiver strain and the quality of family life

- The quality of the relationship between the clinician and the child/family
- The child's readiness to change
- Consumer satisfaction with the clinical process
- The goals of treatment
- The modality, strategy and tactics of treatment
- The adherence of the child and family to the treatment plan

We will present information from stakeholders concerning which of these domains they consider most important (see Chapter III).

Flexibility: A measurement system should be flexible. Service delivery organizations are at different levels of development and vary in their receptiveness to measurement. Furthermore, the premature implementation of a mandatory system would impede experimentation with alternative measurement systems. A modular system would allow organizations to select those modules they consider most appropriate to their local needs. Modules will differ according to domain, timing of data collection, and respondent. Nevertheless, each module should be standardised, with the potential to contribute to a national database and policy decision making.

Potential for improving clinical effectiveness: A measurement system should provide information that enhances clinical effectiveness. The system should allow concurrent monitoring and encompass service-level variables such as *satisfaction with services* and the *quality of the therapeutic relationship*. Service-level information may help clinicians to maximise treatment effectiveness. However, without definable clinical procedures, outcome data alone are unlikely to enhance services. This issue is discussed at greater length later in this report (see Chapter V: The Modality, Strategy, Tactics, Timing and Dosage of Treatment.

Psychometric soundness: Regardless of its comprehensiveness and feasibility, the measurement system must be psychometrically sound. That is, it must be reliable, valid and sensitive to the kind of change engendered by treatment. Chapter IV reviews in detail the psychometric qualities of 178 candidate instruments, and describes the best 31 instruments in detail.

Developmental sensitivity: Normal development must be understood if developmental delay is to be diagnosed. The pattern and intensity of behaviour that indicates psychopathology differs at different developmental stages. Furthermore, much remains to be learnt about the continuity and discontinuity of psychological disorders across the life span.

Cultural sensitivity: The population served in child and adolescent mental health is ethnically diverse. Values and beliefs embedded in cultural identity can affect a family's interaction with the mental health system. For example, culture influences the following important factors: the understanding of the nature of mental illness; the perceived stigma of attending formal mental health services; the way in which illness concerns are expressed; the way the family conceive of treatment; the relevance of traditional healing methods; and particular perspectives on family burden, responsibility and satisfaction (Snowden, 1996). A measurement system should be sensitive to these issues. For example, a measurement instrument can be translated into different languages; focus groups can provide guidance concerning such matters as the wording of proposed items and the relevance of items to participants (Hughes & Dumont, 1993); and protocol analysis can provide information about how participants react and respond to the questions they are asked (Hines, 1993). A measurement system designed for use in Australia should have norms and cut-off points that are derived from a representative Australian sample.

Issues Related to Implementation

Is a Good Consumer-Level Outcome System Sufficient to Manage and Improve Services?

Outcome data may be helpful to management yet not serve consumers or clinicians very well. Trauer (1998) contends that measures of organizational process and measures of outcome should be distinguished. Bickman (1997) maintains that outcome measurement alone is insufficient to monitor and improve mental health services. Measurement must be connected to decision making. We will analyse and discuss the need for a system that evaluates clinical processes in such a way that consumer outcome data are related to clinical, managerial, and policy decision making. We will also describe how outcome data could become the engine of continuous quality improvement (see appendix 7).

What Ethical Issues Are Raised by a Measurement System?

Of primary importance is the protection of privacy and confidentiality, particularly in a system that provides information for centralised decision making. Furthermore, there is a serious danger that invalid measurement could lead to false diagnosis, inappropriate treatment, premature or excessively delayed termination of treatment, and erroneous policy decisions.

How Should an Outcome System be Implemented in Australia?

Even if the measurement system met all design criteria, it might still not be successful unless the processes of implementation were carefully considered. The imposition of a measurement system without consultation and cooperation is likely to alienate frontline clinicians to the extent that they refuse to use the system or even sabotage it. We will consider how to enlist practitioners and managers in the design, installation and maintenance of such a system. Attention must also be paid to the resources (personnel, equipment, training, expenses, and installation time) required to develop, refine and maintain the system. Maintenance is especially important because the system will require continual updating. The type of staff needed to install and supervise the operation of the measurement system, its computerization, and the training required to operate it are important considerations. Above all, a climate favourable to measurement and self-examination must be engendered.

Borrow, Buy or Build?

Few standardised instruments have been internationally accepted as process or outcome measures; on the other hand, the building of a new system takes time and money. In Chapters IV and V of this report, we will analyse and discuss the advantages and disadvantages of adopting or buying existing instruments in contrast to developing an Australian system de novo.

Summary

In principle, a standardised *baseline-followup and concurrent measurement system* is designed for the purpose of improving services. A measurement system is the foundation of continuous quality improvement. Outcome measures should be multifaceted, comprehensive, derived from several sources, feasible, psychometrically sound, developmentally sensitive, and culturally informed.

Information collected at baseline and progressively at well-chosen times should be conveyed in timely, predictable fashion, in a form useful for decision making, to the key decision makers. The system should be adaptable to all levels of care and sufficiently flexible to allow different mental health services to choose the elements of the system most appropriate to their needs and resources. A modular system would provide the flexibility required.

By basing their reasoning on accurate information about the process and outcome of services, rather than on compliance-oriented data, clinicians, managers, consumers. administrators and policy makers will move from rule-driven decision making to decisions based on practical information. Furthermore, reliable information, conveyed in a timely, predictable fashion, will provide the basis for continuous quality improvement. Objective measurement is the means by which all those with a legitimate interest in mental health services, and those responsible for the improvement of those services, can make rational decisions.

This consultancy solicited stakeholders' perspectives on outcome measurement, reviewed existing instrumentation, and recommends a strategy for implementing a measurement system in Australia. It is apparent that no single measure can satisfy the requirements of a comprehensive system. A composite of measures is needed. However, it is unclear whether the final product should be composed of separate instruments derived from the public domain, or if it should be assembled from instruments purchased from commercial sources, or whether it must be designed and developed afresh, wholly or in part, as an Australian system. These questions are considered further in Chapter V.

A glossary of terms related to outcome measurement is provided in Appendix 1.

CHAPTER III

CONSULTING THE STAKEHOLDERS

A Review of the Literature Concerning the Implementation of Outcome Measurement

Barriers to Design and Implementation

Stedman et al. (1997) state, “If a measure is meant to be used by service providers and consumers to inform treatment progress, then it is only reasonable to expect that consumers and service providers will have a great deal of input into the design and implementation of such a system”. The development of comprehensive concurrent and baseline & follow-up measurement systems should not be embarked upon without careful attention to the perspectives of stakeholders. However, only limited information is available on this subject (Chappell & Branch, 1993; Smith, Fischer, Nordquist, Mosley, & Ledbetter, 1997; Stedman et al., 1997). While two Australian studies (Andrews et al., 1994; Stedman et al., 1997) have addressed stakeholder perspectives concerning the design and implementation of measurement systems in adult mental health, there is a dearth of national or international literature concerning outcome measurement in child and adolescent mental health.

A crucial consideration for any review of implementation of measurement systems in child and adolescent mental health is the question of whether consumers, carers, clinicians, managers and policy makers support the concept. Andrews et al. and Stedman et al. conclude that stakeholders support outcome measurement in theory but have significant concerns about how it will be implemented. While there has been no systematic assessment of the perspectives of practising clinicians, the literature suggests that clinicians are generally not interested in outcome measurement or somewhat hostile, viewing it as primarily a research endeavour (Eisen & Dickey, 1996).

Clinicians may be resistant to the introduction of outcome measurement because they suspect the system will challenge their autonomy and expertise or that it will be used as a tool to cut costs. Clinician resistance can also be anticipated if clinicians do not understand the measurement system, are unable to utilise the information provided, or find the system burdensome. Moreover, no systematic study has shown that data can be used successfully to improve services and clinical outcomes.

Consumers and carers may be resistant because the system is too burdensome or complex, or a threat to confidentiality (Smith, Waxman, Snyder & Raphael, 1996). Managers and policy-makers may be resistant because they are wary of innovative approaches that are potentially costly and disruptive to service delivery, particularly if they regard existing services as satisfactory (Pratt & Moreland, 1996).

The literature emphasises the importance of addressing these resistances if implementation is to succeed. It is argued that stakeholder resistance will be reduced if the utility of the system and its implications for stakeholders are made clear (Dornelas, Correll, Lothstein & Wilber, 1996; Smith, Manderscheid, Flynn, Steinwachs,1997;). Measurement systems should be brief and user-friendly (Andrews et al., 1994), easily understood and clearly formulated (Stedman et al., 1997). Clinicians need short-term outcome feedback information (Bickman et al., in press). Systems should be described in a way consistent with the language of treatment planning, progress review and clinical supervision (Newman, Hunter & Irving, 1987). Minimal burden should be placed on any stakeholder (Smith, Fischer et al., 1997). There is a clear tension between the need for an outcome measurement system that is comprehensive and one that is neither too complicated nor too time-consuming.

It is argued that implementation will be facilitated if modest projects are piloted, paying attention to users' experience of the new system (Andrews et al., 1994; Pfeiffer, 1996). The recruitment of key stakeholders into the implementation process is thought to mitigate barriers. Clinician champions, dedicated project managers, and vigorous researchers are required (Pfeiffer & Shott, 1996; Tobin & Hickie, 1998). Management leadership is crucial (Tobin & Hickie, 1998). Managers can promote outcome measurement as an opportunity for quality improvement, allay clinicians' fears of administrative intrusiveness, and ensure the availability of appropriate resources and incentives (Stedman et al., 1997).

Education and Support

Hernandez and Hodges (1996) consider that an outcome measurement system will not be implemented unless the political climate among stakeholders is favourable, adequate resources are provided, and the commitment of leaders is sustained. Glaser and Backer (1980) assert that a measurement system will fail to take root unless it is congruent with the values of the organisation and perceived as valuable by all stakeholders. There is no specific literature concerning the educational programs required to inform stakeholders about the concept and technical details of outcome measurement, to create a favourable political climate among stakeholders, and to align measurement with the values of organisations. So little is known about these important issues that the educational process will need to involve everybody: researcher-educators, managers, policy-makers, consumers, carers and clinicians. These questions are addressed at greater length in Chapter V (Educating the Stakeholders).

Computerisation

Computer-aided outcome measurement has been developed for assessment, for the tracking of clinical progress, and to provide aggregate data across services. Stedman et al., found that the use of computers for data collection did not affect the results of their study. Newman et al. (1987) assert that “even with a non-automated record system, the administration and scoring can be easily managed if properly integrated and used as an integral part of the clinical assessment, treatment planning and review process.” Computerisation is essential if data are to be integrated at the site level or across sites. Clinician use may also be enhanced if the computerised system provides timely feedback.

The acceptance by stakeholders of computerised systems (Sederer, Hermann & Dickey, 1995), and the impact of this on the success of the outcome measurement system, must be considered. The computer facilities available within services and the cost of introducing computerisation (including education and training) must be considered, along with methods of administration, data storage, and retrieval.

Ethical Issues and Risks

Ethical Issues

Stedman et al. found that the issue of confidentiality was raised on numerous occasions. Stakeholders express concern about who has access to (and might mismanage) client data (Pfeiffer, 1996). Some are concerned that client followup will violate confidentiality. Unless anonymity and confidentiality can be guaranteed, it is unlikely that consumers will consent to, or comply with, outcome measurement (Stedman et al., 1997). Clinicians express concern that outcome measurement will intrude on the therapeutic alliance, and that treatment might be based on invalid measurement data.

Risks

Stakeholders perceive a risk that outcome data could be used by managers to determine the eligibility of clients for treatment. Without sufficient commitment and partnership between management and other stakeholders, measurement systems could become divorced from day-to-day service delivery (Tobin & Hickie, 1998). Above all, stakeholders fear that an outcome measurement system will become an auditing exercise linked to resource allocation rather than a tool for quality improvement.

Summary

The literature suggests that successful implementation requires an understanding of the barriers to implementation. Stakeholder education, training and support are essential. Ethical risks and potentially adverse effects must be acknowledged and avoided. We explored several of these issues with stakeholders. A description of their perspectives follows.

The Methodology of Stakeholder Consultation

The consultancy undertook a national and statewide consultation with consumers, administrators and practitioners in child and adolescent mental health. We used three research techniques: telephone survey, mail survey, and focus group. A semi-structured questionnaire formed the basis of the telephone and mail surveys and guided discussion in the focus groups. The format of the semi-structured questionnaire can be found in Appendix 2. The framework for focus group discussion can be found in Appendix 3.

Senior Mental Health Academics and Practitioners

Thirty-two senior academics, practitioners in the field of child and adolescent mental health, and researchers who had significant experience in outcome measurement, participated in a national telephone survey. Each respondent was asked to recommend other experts in the field who should be consulted. Every new person subsequently contacted was asked to recommend others. Attempts were made to recruit participants from the disciplines of psychiatry, psychology, sociology, and mental health nursing. Liaison with other professionals (e.g., at the March 1998 National Mental Health Outcome Conference in Sydney) and a review of literature provided further contacts, all of whom were pursued by the research team. In all, 38 people were invited to participate and 32 did so.

Procedure: Participants received a letter from the investigators outlining the consultancy brief and inviting them to participate in a semi-structured telephone interview. This letter was followed by a telephone call from the research team to arrange an interview time. Prior to the interview, a copy of the questionnaire was forwarded to participants together with a covering letter that outlined the ethics of the interview. Participants were assured that, without prior agreement between all parties, they would not be identified in any report resulting from the interview. Two members of the research team conducted the telephone interviews. Interviews lasted from 45 to 55 minutes. Interviewers took notes and audiotaped all interviews. Secretarial staff subsequently transcribed the interviews.

Peak Bodies, State Mental Health Directors and Child and Youth Mental Health Centres.

Peak Bodies

Eight peak bodies were invited by letter to respond to the following points either by questionnaire or submission:

- Did members of the organisation use outcome measurement instruments?
- What were the advantages and disadvantages of adopting existing instruments in contrast to developing an Australian system?
- Which domains should be incorporated in a comprehensive system?

- What was the organisation's perspective on the risks and problems of introducing such a system into a clinical setting?

The questionnaire was a slightly amended version of the semi-structured format used in the telephone interviews.

The peak bodies who were invited to respond within at least four weeks, were as follows:

- The Australian Association of Social Workers
- The Australian Medical Association
- The Australian and New Zealand College of Mental Health Nurses
- The Australian Psychological Society
- The National Community Advisory Group
- The Occupational Therapists Association of Queensland
- The Royal Australian College of General Practitioners
- The Royal Australian and New Zealand College of Psychiatrists

Reminder letters were sent to all of the organisations in the third week following our invitation, except to the National Community Advisory Group, which received a follow-up phone call from the research team. Four responses were received, from the following peak bodies: The Australian Psychological Society completed a questionnaire; The Royal Australian and New Zealand College of Psychiatrists made a submission; The Australian and New Zealand College of Mental Health Nurses and the National Community Advisory Group indicated that they were unable to make submissions or complete our questionnaire but provided the Consultancy with brief statements outlining their positions.

State Mental Health Departments

Government Mental Health Services in all States and Territories were invited by letter to participate. Completed questionnaires were received from four State Mental Health Services.

South Australia	• Division of Mental Health, Women's and Children's Hospital.
	• Mental Health Unit, South Australian Health Commission.
Northern Territory	• Darwin Urban Mental Health Service
	• Central Australia Mental Health Services
Tasmania	• Community and Rural Mental Health Services
ACT	• Child and Adolescent Mental Health Program, ACT Mental Health Services

Submissions were received from The Mental Health Unit of Queensland Health, The Centre for Mental Health, NSW Health Department, and The Mental Health Branch of the Victorian Department of Human Services.

CYMHS Queensland

Seventeen child and youth mental health centres, within the state's 16 district health services were contacted. Sixteen survey documents were returned, representing responses from 14 of the 16 district health services.

Focus Groups

Focus groups are designed to obtain specified information from particular people. The format involves small group discussion on a particular topic under the direction of a moderator. The role of the moderator is to promote interaction, probe for details when warranted, and ensure that discussion remains directed toward the topic of interest (Stewart & Shamdasani, 1990). One strength of the focus group is its ability to clarify participants' understanding of a topic of interest and to elicit their experiences and perspectives (Packer, Race & Hotch, 1994). The open format of the focus group provides an opportunity to elicit information while reacting to and building upon the responses of group members (Stewart & Shamdasani, 1990).

However, focus groups have limitations. Results can be difficult to interpret due to the open-ended nature of responses obtained (Knodel, 1993). The moderator can bias results by knowingly or unknowingly providing cues about the type of responses that are desirable. Furthermore, the small number of respondents and the recruitment process of most focus groups significantly limits the generalisability of the results (Stewart & Shamdasani, 1990).

Consumers

Young people and parents or carers of young persons who use mental health services were consulted through focus group discussions and a survey that aimed to assess parental views on outcome measurement.

Adolescents

Two focus group discussions were conducted with 14 adolescents from the North Brisbane and Caboolture Child and Youth Mental Health Services. Team leaders from these services asked suitable young people, who had accessed their services, whether they might be willing to participate in a group discussion about outcome measurement. Four participants attended the first focus group held in Brisbane. Ten attended the second group in Caboolture CYMHS.

Parents

A focus group discussion was held with six parents or carers of adolescents who had used mental health services. Team leaders of the four North Brisbane CYMHS were asked for the names of carers who might be willing to participate in a group discussion about outcome measurement. The six parents referred were contacted by the research team. After receiving an outline of the purpose and topic of discussion they were invited to attend a focus group. All were willing to participate, but only three could attend on the day and time scheduled. The three other participants who attended were recruited from a CYMHS Carer Advisory Group, made up of five parents of clients who use these services. All members of this group were invited to attend. Two were unable to do so.

Further consultation with parents was undertaken by means of a survey. Questionnaires assessing parental attitudes to outcome measurement were distributed to the four CYMHS clinics in the North Brisbane region. The questionnaire (Appendix 4) consisted of 6 questions with different sub-items rated on a 5-point scale ranging from 1(not at all) to 5 (extremely). In all, 40 parents completed this questionnaire.

Clinicians

Consultation with clinicians was limited to employees of the Brisbane Royal Children's Hospital and District Child and Youth Mental Health Service that operates three community clinics and one inpatient unit. All 84 clinical staff was invited to participate in the consultation process either through focus group discussion or by completing the questionnaire. Clinicians participated in focus groups according to their discipline. Consequently, focus groups were conducted with psychologists, social workers, speech pathologists, medical officers and nurses, involving a total of 43 clinicians, as follows:

Psychologists	9	Medical Officers	9
Social Workers	8	Speech Pathologists	9
Nurses	8		

Thirteen consultant psychiatrists and senior professionals within these services were asked to respond to the semi-structured questionnaire. Ten completed questionnaires were received.

Procedure: Two members of the research team facilitated the discussion with the professional and carer focus groups, while team members from Caboolture CYMHS and Nundah CYMHS respectively co-facilitated the adolescent consumer focus groups with our research team. All discussion groups followed a similar framework, beginning with an introduction and a discussion of participants' previous experience with outcome measurement, and continuing with an exploration of the definition of outcome measurement, the domains to be covered, and the methodology of data collection. Other areas explored included the use and applicability of outcome data, feasibility, support, and the implementation of outcome measurement. Moderators adapted the discussion format for both carer and adolescent groups. All groups took about one and a half-hours. Discussions were audiotaped and recorded by note taking. Tape recordings were later transcribed to facilitate data analysis.

Private Health Insurers

Representatives of private health insurers were contacted. It was agreed that a meeting between the consultancy staff and a panel of representatives of the industry would be more suitable than administering questionnaires or holding a focus group discussion. Two senior members of the consultancy team and a panel of industry representatives met to discuss the insurers' perspective.

Analysis of Qualitative Data

We obtained a detailed, systematic record of the themes and issues arising from the interviews and questionnaires. Domain analysis (Neuman, 1994) was applied to these qualitative data. Data were grouped according to nine domains derived from the key research questions. Important themes and concepts were identified, and their relationship to each other examined. The significance of data was established through qualitative analysis and interpretation as well as frequency counts of responses. The nine domains of data analysis were as follows: -

- **Participants' profile:** Previous or current experience in outcome measurement
- **Definition:** The ideal characteristics and breadth of coverage of an outcome measurement system
- **Domains:** The areas of measurement to be covered and the methodology of measurement including frequency, means of data collection, and sources of information
- **Applicability:** The usefulness of the outcome information to clinicians, services and policy makers
- **Existing systems:** Outcome measurement systems currently in routine clinical use
- **Feasibility:** The perceived importance of outcome measurement to child and adolescent mental health, and the degree of staff support and commitment to outcome measurement
- **Implementation:** The barriers to implementing outcome measurement, its risks and adverse effects, and the resources available
- **Recommendations:** Strategies for overcoming barriers, and the education, training and computerisation required for implementation of the measurement system.

A description of stakeholder groups approached, their response rates, and the number of eventual participants is provided in Table III/1.

Table III / 1—Stakeholder Consultation: Groups Consulted and Response Rates

Group	Method	No. invited to participate or who were sent questionnaires	No. of participants or completed interviews/ questionnaires or submissions received	Return rate in percentage
Senior Academics and Practitioners	Telephone interview	38 (32 confirmed)	32	100
State Mental Health Services	Semi-Structured Questionnaire or submission	8	7	87.5
Peak Bodies	Semi-Structured Questionnaire or submission	8	4	50
CYMHS (QLD)	Semi-Structured Questionnaire	17	16	94
North Brisbane CYMHS Child Psychiatrists/Senior Professionals	Semi-Structured Questionnaire	13	10	77
Parents/Carers	Attitudes to Outcome Measurement Questionnaire	(distributed to parents/carers over 3-week period in North Brisbane CYMHS)	40	-
	Focus Group	6	-	-
Adolescents	Focus Group		14	-
Clinicians	Focus Group	77	43	56

Stakeholder Perspectives on Outcome Measurement

This section describes the results of the data-gathering activities. It should be noted that not all respondents were well informed about the complexities of outcome measurement. Indeed, many acknowledged their limited knowledge. Thus, our findings reflect stakeholder perspectives from widely varying experiential and knowledge bases. It is important that these findings be balanced against the opinions of those with expertise in the field.

The Definition of a Measurement System

Clinicians found it difficult to define "measurement system". It was apparent that some stakeholders had not entertained the concept of a "system". Others had thought about the concept but were uncertain of how it should be defined. To the extent they can define it, stakeholders regard an outcome measurement system as made up of objective, standardised measures that tap a range of specified domains before and after intervention. These measures should utilise data from multiple informants. Measures of consumer satisfaction are essential. An outcome measurement system assesses positive or negative client change following intervention, gauging the short and long-term consequences of treatment at three levels: *individual patient; patient groups;* and *services*.

The Importance of Outcome Measurement

Stakeholder opinions concerning the importance of a comprehensive, concurrent and baseline-followup measurement system vary considerably. The majority of consumers, carers, senior academics and professionals, state directors and national bodies strongly believe that such a system is important. Strong carer support was registered, all items dealing with measurement being rated above 4 (very), on a 1-5 rating scale. Many stakeholders see a measurement system as providing important data at clinical, service and systems levels, locally, at state level, and nationally.

In contrast, clinicians have mixed feelings about the importance of a measurement system. Some believe that, depending on the integrity of the system and its application, *concurrent* and *baseline-followup* measurement could be important to child and adolescent mental health. Others consider that the potential importance of a measurement system is counterbalanced by the danger that such a system would "mechanise and measure everything", draining already limited resources from clients and clinical work. A minority of stakeholders in most stakeholder groups are hostile to a measurement system. These respondents are concerned about the burden of the system and its potential restrictiveness and intrusiveness in relation to clinical practice. While the majority of respondents believe an outcome measurement system is important, a significant number qualify their responses, seeing importance as contingent upon how well the system is designed, implemented and utilised.

The Utility of Outcome Measurement

Stakeholders believe that a concurrent and baseline-follow up measurement system has a number of potential uses. Most anticipate that sophisticated measurement will yield a greater understanding of clinical process and outcome, with the potential to enhance service effectiveness and facilitate rational, equitable resource allocation. Data from a measurement system could validate clinical work and enhance the credibility of child and adolescent mental health in the wider health community, raising morale, generating energy, and stimulating research.

The majority of stakeholders think that a measurement system could help clinicians treat clients. Many regard measurement as an essential part of therapy. Outcome measurement systems are regarded as tools to help professionals identify and conceptualise problems, review their work with clients, make evidence-based decisions in regard to treatment, and gauge the effectiveness of treatment.

The Use of a Measurement System to Assess Quality

Outcome measurement is seen as an important tool in continuous quality improvement ("part of the quality loop"). In the words of one respondent, "If we don't measure our treatment, we do not know where we go, to know whether what we do is effective or not." Outcome measurement could also promote a common language and common standards, through which clients and clinicians could establish goals, review treatment progress, and determine outcome.

Carers consider that an outcome measurement system is important because it would increase clinician and service accountability, assisting services to target the needs of parents and clients. ("Something needs to be done. Parents need something concrete.")

Clinicians generally agree that a concurrent and baseline-followup measurement system could help clinicians treat clients, but that this will be dependent on whether the measurement system is sufficiently sophisticated to address a range of disorders whilst at the same time remaining feasible.

Concerns were raised by a number of stakeholders that the usefulness of a measurement system would depend on whether clinicians are involved in its development and implementation. If not, clinicians would not use the system.

Stakeholders hold unequivocally that, if the measurement system is valid, sophisticated, and capable of reflecting case complexity, it should be used by organisations to measure the quality of their services and to ensure the accountability of those services. In this regard, some respondents fear that outcome measurement alone does not adequately reflect service quality. Administrative funding decisions based solely on outcome measurement could be inaccurate and inequitable. Clinicians could become negatively biased in the way they utilise the system. These stakeholders stress that outcome measurement should not be the sole means by which a service is judged.

While supporting the use of a measurement system to measure quality, a minority was sceptical about the possibility of designing an adequate system for the wide range of child and adolescent mental health problems.

The Advisability of Adopting a Standard System

Most stakeholders alluded to the diversity of services and types of treatment in child and adolescent mental health services. Some see this diversity as a potential barrier to the introduction of a standard measure. Others recommend the provision of optional or discretionary measures which individual services could select, according to their needs. Most stakeholders seek assurance that a universal measure or "endorsed battery" should be well validated. Stakeholder opinions fall into the following three categories:

- A core set of measures should be used throughout Australia. This would enable consumers and clinics to be compared against national figures. As a supplement to the core measures, there should be "add-on", or "modular" measures designed for particular conditions and treatments.
- There should be a single, standard measure. One respondent was concerned that such a measure should not be "set in concrete". Another stated that, while very desirable, a standard measure would be difficult to establish.

- A third group of stakeholders oppose standardisation, which they consider unfeasible or unrealistic. Instead, they see a need for “a battery or a selection of tools that are recognised, recommended, accessible, and available”. Concerns were raised that, if a single standard measure were imposed, services and clinicians would lose self-determination and control, and that the diversity of client groups across Australia would be obscured.

The Use of a Measurement System by the Federal Government to Evaluate Cost-Effectiveness

Most respondents are concerned that the federal government might use a measurement system to evaluate the cost-effectiveness of services. A number of respondents oppose the idea because the information could be abused by decision-makers who either had particular political agendas or who did not understand the complexity of child and adolescent mental health. In the words of one stakeholder, “I sure wouldn’t want to give that over to some power-hungry minister who’s just looking for some way to advance his career and misuses that information. There’s tremendous potential for abuse”.

Respondents express concern that outcome evaluation might fail to take into account the often-extended course of mental health treatment, and that there might be a tendency to link cost-effectiveness with short-term gains. Furthermore, if the measurement system were used to assess cost-effectiveness and determine resource allocation, it could lead to falsification of data by clinicians and services, a preference for “the easiest to treat”, and the misuse of outcome data by government to “squeeze services” by withholding funding. The director of a rural mental health service reported that, if treatment outcome is translated directly to the costing of services, it would be extremely difficult for small or rural services to compete with larger services that have more resources. Many stakeholders doubted that any available measurement system is sophisticated enough to evaluate cost-effectiveness accurately.

The Ideal Measurement System

Respondents believe that a measurement system should maximise the use of data collected in routine clinical practice and be *congruent with established treatment planning and review. Multiple informants* should be involved and *multiple outcomes* measured. The system should be *brief, elegant, and simple to* *complete, score and interpret*. Clients and parents want a system that is *non-intimidating* and "*user-friendly*". The system should not dominate the treatment process and it should provide clear, *comprehensible feedback* to all stakeholders. Stakeholders stress that an ideal measurement system should only collect data that are *clinically relevant*. One respondent summarised the matter as follows: "I would like to see an outcome measurement system that is evidence- based (in which) we're collecting data because we know they are critically related to the causes and the remedies of mental health problems in children and adolescents". The most important criteria for such a system are *feasibility* and *reliability*. Other important criteria are *sensitivity to developmental change*, *sensitivity to change over time* (including long-term change), *psychometric soundness*, *cultural sensitivity* (particularly in relation to the needs of Aborigines, Torres Strait Islanders and people from non-English-speaking backgrounds), *potential to improve clinical effectiveness*, and *sensitivity to the literacy levels of clients*.

Stakeholders agree, on the whole, that the ideal system should encompass the following:

- Social network information
- Intrapsychic factors
- Experiential factors
- Observational data
- Contextual data
- Information about the clinical presentation
- Functional status information
- Quantitative and categorical diagnostic information
- "Whole of life" measures (e.g., concerning education and juvenile justice)

The Comprehensiveness of the System

Most respondents understand that the comprehensiveness of the measurement system is limited by the degree to which it is feasible and practicable.

Comprehensiveness should be traded against brevity and ease of administration. ("If that balance is not achieved the system might be useless to all, as it would be too cumbersome or too simple".)

A minority of stakeholders consider that the system should be uncompromisingly comprehensive because the problems experienced by children and young people with mental health problems are complex.

A significant minority would resolve the tension between comprehensiveness and feasibility by targeting measures. They recommend that the outcome system should not attempt to measure every aspect of a client's progress and outcome but, rather, to relate measurement to treatment goals. ("You have to tailor your outcomes to your objective.")

Domains

What domains or areas should a measurement system address? Stakeholders were asked to nominate the areas of measurement they believe are important. Professional stakeholders were asked to respond to a questionnaire, according to a survey technique adapted from Bickman, Rosof et al (in press). In the original mail survey, 539 United States clinicians were asked to rate 29 different domains. The same domains were presented to our stakeholders who were asked simply whether a measurement system should encompass each domain. Table III:2 compares the way Australian and American respondents rank-ordered the domains.

Australian and United States clinicians agreed that the following domains are highly important:

- Family functioning
- Parent-youth relationship
- Specific areas of functioning

The Australian stakeholders ranked the following domains highly:

- Specific symptom scores
- Burden of care

Table III.2 — RANK ORDER OF IMPORTANCE OF DOMAINS (From most to least important)

		Australia			US
	Survey Items	**All* n = 74**	**Clinicians n = 35**	**Senior Academics and Professionals n = 32**	**US Clinicians n = 539**
1.	Family functioning (e.g., family relations, family conflict, communication among family members)	1	1.5	4	3
2.	Quality of parent-youth relationship	2.5	1.5	5	5
3.	Areas specific to youth's level of functioning (e.g., school, peers, family, community)	2.5	4.5	2	5
4.	Symptom scores for specific disorders/problem areas for youths (e.g., depression, anxiety, drug/ alcohol use, eating disorders, suicidal ideation)	4.5	11	2	19
5.	Parental stress and strain of caring for youth with mental health problems	4.5	7	7	23.5
6.	Youth's physical health and current medications	6	3	9	12.5
7.	Youth's compliance with treatment plan (e.g., completion of therapy "homework" assignments)	7	4.5	11	7.5
8.	Parent's physical and mental health, behavioural symptoms and history with psychotropic medications1	8	6	11	12.5
9.	Parent's compliance with treatment plan (e.g., follow through with disciplinary actions, assisting youth in meeting weekly therapy goals)	9	9.5	11	5
10.	Global score for the youth's problems (e.g., global symptom severity score)	10	19.5	2	29
11.	Youth's global level of functioning score (e.g., functioning score aggregated across domains: school, work, peer groups, family, community)	11	12.5	6	28
12.	Youth's skills and strengths (e.g., social skills, hobbies, and other positive activities)	12	8	17.5	19
13.	Youth's satisfaction with therapist, clinic, and service system.	13	14.5	8	12.5
14.	History of child maltreatment (i.e., physical abuse, sexual abuse, emotional abuse, and neglect)	14.5	9.5	23	1.5
15.	Narrative (open ended) responses from parents or youths on various topics including: medications, goal attainment, or progress (e.g., How effective are the medications?, Do you think you are achieving your goals from therapy?, Overall, are your child's emotional and/or behavioural problems improving, staying the same or getting worse?)	14.5	14.5	13.5	19

Table III.2 — RANK ORDER OF IMPORTANCE OF DOMAINS (From most to least important) CONT'D

		Australia			US
	Survey Items	**All* n = 74**	**Clinicians n = 35**	**Senior Academics and Professionals n = 32**	**US Clinicians n = 539**
16.	Youth involvement in developing treatment plan	16.5	16.5	16	19
17.	Parent involvement in developing treatment plan	16.5	18	13.5	19
18.	Past and present youth stressors (e.g., death of a parent, poor grades, poor relations with family and friends)	18.5	12.5	24	1.5
19.	Therapeutic alliance (e.g., parent's or youth's perception of their relationship with the clinician)	18.5	16.5	19	5
20.	Past and present family stress (e.g., job loss, family member married, childbirth or parent died)	20	21.5	21	7.5
21.	Barriers to care (e.g., financial, transportation, school schedule, parent/teen's work schedule)	21	23	17.5	23.5
22.	Youth's motivation to change	22	21.5	25	12.5
23.	Concrete family resources (e.g., income, health insurance, housing, food)	23	27.5	15	23.5
24.	Youth's past utilisation of mental health services, drug/alcohol treatment, and medical services	24	25	21	23.5
25.	Parent's reasons for seeking services for youth	25	26	21	12.5
26.	Parental attitudes about seeking therapeutic services for youth (e.g., optimistic about potential benefits of services, dislike for mental health professionals)	26.5	24	26	16
27.	Parental expectations for youth with respect to current and future personal and career goals	26.5	19.5	28	26.5
28.	Youth 's attitude about seeking mental health services (e.g., optimistic about potential benefits of services, dislike of mental health professionals)	28	27.5	27	26.5
29.	Youth's reasons for seeking therapeutic services	29	29	29	12.5

* 'All' include clinicians (n=35), state directors (n=7), senior academics and professionals (n=32)

- The youth's health
- The youth's compliance with treatment
- The parent's compliance with treatment
- Global symptom severity score
- The strengths of the youth

In contrast, the American clinicians rated the following domains highly:

- History of maltreatment
- Youth stressors
- Therapeutic alliance
- Family stressors

What can we conclude from these results? First, there is consensus on domains related to *youth functioning*, *family functioning* and *symptoms* as outcomes. These could be the core areas to cover in a *baseline-followup subsystem*. There is consensus on the desirability of data such as adherence to treatment and therapeutic alliance. These are important domains in a concurrent measurement subsystem.

Our stakeholder consultations provided additional potential areas of measurement: *The side effects of treatment* and *the child's emotional state and behaviour*. Consumers stated that *self-esteem* and *happiness* were important to them. Asked to rate measurement domains, carers rated most of the measurement domains highly, all domains being seen as equally important. Stakeholders raised the issue that an outcome measurement system should provide information about *unplanned discharges*. The impact of treatment on the child's or young person's *vulnerability to risk* was also mentioned. The *strength of social networks* and *community resources* available to the consumer and their families were seen as important. One stakeholder mentioned *hours of work* and *hours of care*. Stakeholders considered that measures should be designed for the *burden of care for clients and their families, the accessibility of services*, and *training, supervision and staff development*.

Feasibility

The receptiveness of clinicians to using outcome measurement data varies. A minority are certain that clinicians would use data from a measurement system. At the other end of the spectrum, a small group of respondents stated that they would not use outcome data. The remaining respondents stated that, if certain conditions applied, clinicians would be receptive to using data from an outcome measurement system. Respondents considered that clinicians would utilise a system that focuses on “the really significant issues that affect the patient’s life”. *Functional impairment, symptom reduction,* and *family functioning* are regarded as the key domains. Other variables affecting receptiveness are the burdensomeness of the system, the education of clinicians in how to use the system, and the sophistication of the manner in which data are presented. One set of respondents is of the opinion that clinicians already know how to use outcome data. Others consider that training is required.

Respondents regard clinicians as having limited time to collect and review data. Most respondents consider that clinicians could devote no more than five minutes per session. A number of respondents advise that data collection and review should be incorporated routinely into clinical practice and not treated as a separate, time-specific procedure. Some clinicians already collect data at every session and “build it into the transactions between the clinician and the client”. Other respondents considered that data collection and review should not occur on a sessional basis, but rather at intake, discharge and followup. Stakeholders consider that the amount of time clinicians can devote to data collection and review relates to the simplicity of the data and its ease of its collection.

Consumers and parents are seen by some as both more willing to devote time and to have less self-determination about whether they do so. (“Clients often don’t have a choice. The condition of therapy is that they complete data collection.”) Respondents see parents and consumers as having more time to provide data than clinicians. Opinions concerning the time available varied from 10 to 20 minutes. Consumers and parents are thought to accept the need to provide data, particularly in the early stages of treatment. Consumers said that the time they would devote to data collection would depend on whether the experience was positive (e.g., non-repetitious).

Procedures for Data Collection

Stakeholders were asked how concurrent and outcome data should be gathered, how frequently data should be collected, and who should collect the data.

The Method of Data Collection

A number of stakeholders believe that the method of data collection should be determined by the stage of treatment, the type of information required, and the individual client.

Stakeholders support the collection of data through interviews, either face-to-face or by telephone. The interviewer should be supportive and helpful. While stakeholders (aside from consumers) do not object in principle to mail-out interviews, most commented that return rates would be poor. Stakeholders consider that data collection should be professionally undertaken, questionnaires well and attractively presented, and administration “tied to the normal operation of the clinic”.

Stakeholders are mixed in their attitude to computerised data collection (e.g., using touch-screen technology). Some think that clients would accept computer data collection whilst others predict that parents or children would be non-compliant. In general, carers were more positive towards the collection of data from face-to-face interviews and questionnaires than from a computer. A number of stakeholders warned that universal computerisation cannot be assumed.

Consumers differ in how they would like to provide data. Some young people would prefer a face-to-face interview, others a questionnaire. Some would prefer to provide unstructured feedback (e.g., through paintings, drawings, poetry, or diary entries). Young consumers said that they would like to have the option of providing data either in their homes or in clinics at weekends.

The Frequency of Data Collection

Consumers and parents consider that data should be collected at intake and at the end of treatment. A significant proportion is opposed to concurrent data collection. Follow-up data collection (up to a maximum of six months post-termination) is regarded as crucial. Several consumers consider follow-up interviews to be therapeutic.

State directors, child psychiatrists, senior professionals, and parents support regular data collection at intake, concurrently with treatment, at termination, and up to twelve months after the commencement of treatment. The majority of stakeholders do not recommend follow-up after twelve months. Psychologists suggested that concurrent data collection should be integrated into the clinical process to ensure greater compliance.

Adolescents would prefer to provide information at those times when they feel it would be appropriate to report on their progress. A number of young people fear that regular data collection would place demands on them at times when they do not wish to communicate or have little to say. They do not regard regular data collection as important or appropriate. Nurses and young people had reservations about collecting data concurrently with treatment. Nurses were concerned that concurrent measurement could endanger the therapeutic alliance.

The Sources of Data

With the exception of consumers, stakeholders unanimously agree that informants should be young people, parents or carers, teachers, and mental health clinicians. Clinicians vary in their opinion as to the minimum age at which a child can supply appropriate data. Psychologists suggested that information from children below 12 years of age should be collected by means of an interactive assessment. Some stakeholders suggested that guidance officers, general practitioners, and agencies responsible for the care and protection of young people, could also be informants.

Consumers have mixed feelings about collecting information from teachers and parents. Consumers believe that parents can provide valuable information regarding changes in the young person's mental state, the home situation, and family dynamics; however, the usefulness of such information is thought to depend upon the parents' involvement with the young person and their commitment to the young person's treatment. Consumers were adamant that parents should be involved only if the youth gives permission. Young people are quite ambivalent about teacher involvement. While teachers are seen as having the capacity to provide useful information, most consumers feel uncomfortable at the thought of a teacher filling out a questionnaire about them, particularly if they did not have a good relationship with the teacher.

Some consumers expressed mistrust of teachers, fearing that teachers would publicly humiliate them if asked by a mental health service to provide data.

Borrow, Buy Or Build?

Several stakeholders did not respond to this question or considered they did not have sufficient knowledge to guide the consultancy in this matter. Many were unaware of any current outcome measurement system that could be used in child and adolescent mental health. Stakeholders, who were aware of an existing system, were often unable to describe it.

Stakeholders are divided about whether an outcome measurement system should be adopted from elsewhere or designed afresh. Policy makers and telephone survey respondents strongly recommend a systematic review of existing measurement instruments. The most frequently cited instruments were the Child Behaviour Checklist and Youth Self Report, the HoNOSCA, and the Conners Rating Scales. These stakeholders see the borrow/buy approach as cost-effective and optimising existing research. ("To build from scratch is a bit arrogant." "Starting from scratch will be a ten-year, ten million dollar exercise.")

Some clinician stakeholders, from all disciplines within child and adolescent mental health, are supportive of designing a new system, however they provided little supporting evidence for this position. Several stakeholders are excited about the possibility of developing an Australian system.

Commitment and Support

While there is general endorsement of the importance of outcome measurement most stakeholders believe that support for a measurement system is likely to be affected by the following factors:

- Clinician resistance.
- The burden of administration and data review.
- The role of stakeholders in the design and development of the system.
- Cost.
- Usefulness.

- Comprehensibility
- The impact of the system on clients

Respondents agreed about the levels of investment and commitment that could be expected within different professional groups.

Researchers: Stakeholders are unanimous in their belief that researchers will be prepared to make a significant commitment and investment. Respondents see outcome measurement as “primarily an area of research at this time”. There is a suggestion that researchers’ enthusiasm will be motivated by the research opportunities they perceive such a system as affording.

Administrators: High administrator investment and commitment are anticipated, motivated by the desire for tools to assist in the management of costs.

Mental Health Clinicians: Mental health clinicians are expected to be less enthusiastic than other groups. Stakeholders predict that clinician support will be limited, primarily because of workload pressure and perceived lack of organisational support to clinicians in implementing the system. Social workers and medical officers said that they had little support for such a system. With some exceptions, medical officers expressed scepticism that any system could define outcome. Social workers were particularly concerned about what they perceive as the risks of a measurement system. Both groups stated that they had little time to invest in measurement. Other clinical groups professed qualified support, provided they were satisfied with the system’s design, validity, feasibility, usefulness, and integration into clinic practice.

Clients, Parents and Caregivers: Stakeholder groups gauge parent and consumer support as “moderate” to “significant”. Clients think that the characteristics of the system will affect their support; for example, if it is time-consuming, “boring, stupid or repetitive”. Clients believe their parents’ investment will depend on whether the parent initiated treatment and how committed they are to their child. Parents say they would support an outcome system if it is “well-organised, sensitive to parents’ needs, user-friendly and non-bureaucratic”. Other stakeholder groups consider that client and parent commitment and support depends upon the following:

- The characteristics of the parents and clients (e.g., socio-economic status)
- Parents' and clients' compliance with, and involvement in, treatment
- How well the system is understood by clients and parents
- Whether the data is integrated into routine clinical practice
- Client and parent involvement in system design and implementation
- The enthusiasm (or otherwise) of the clinician for the system
- Time requirement (clients and their parents are thought to be more tolerant than clinicians of time demands)

Sustained client and parent support after discharge is not expected. Long-term follow-up is regarded as problematic.

Private Health Insurers: Representatives of the private health insurance industry express enthusiasm for the development and implementation of concurrent and baseline-followup measurement systems in child and adolescent mental health. Historically, mental health services have been a costly area for health insurers. Legislative requirements and clinical practice can affect the cost of mental health services. Insurers want better information concerning the nature of mental health service delivery and the effectiveness of mental health services. They express concern that they pay for services the nature and effectiveness of which is unclear to them. As a result, they find it difficult to determine which programmes they should fund.

Insurers state that an outcome measurement system is needed for the following purposes:

- To provide a common language amongst stakeholders
- To clarify clinical intervention
- To provide a legitimate standardised scale for the measurement of client progress and outcome
- To add rigour to mental health services.

Insurers support the need for multiple informants, including the consumer, carer and clinician. Multiple domains of measurement are also regarded as essential, and functioning as particularly important.

Insurers would like to guarantee that their customers are receiving high quality and effective services, but they assert that they could not defend the costs of these services without outcome data. While there is understanding and support of the need for careful development of an outcome measurement system, insurers want a system to be developed as quickly as possible.

Peak Associations: Eight peak associations were invited to submit their views regarding Outcome Measurement in Child and Adolescent Mental Health, via a structured questionnaire or a written submission (see Appendix 2 for the pro-forma). Submissions or statements were received from the Australian Psychological Society, the Australian and New Zealand College of Mental Health Nurses, The National Community Advisory Group, The Royal Australian and New Zealand College of Psychiatrists.

The Australian Psychological Society stated that a comprehensive measurement system was very important to child and adolescent mental health. The Association would encourage the adoption of such a system. Features of the system seen by the Association as important include the following:

- Multiple domains
- Multiple informants
- Relevance to clinicians
- Brevity

An outcome measurement system was regarded as being useful to psychologists in areas of assessment, treatment and evaluation of treatment. The Association saw no significant barriers to implementation of a measurement system if it were perceived as useful, relevant, and not burdensome. The Association supported computerisation. An outcome measurement system could be used to evaluate quality and the cost of services.

The Australian and New Zealand College of Mental Health Nurses was unable to provide a detailed response. However, the College expressed support for the consultancy and commented that appropriate systems need development.

Telephone discussion and written communication with the National Community Advisory Group indicated that this body was not in a position to comment about outcome measurement in child and adolescent mental health but that it supported “the convening of a number of specific focus groups to acquire consumer and carer input”.

The Royal Australian and New Zealand College of Psychiatrists addressed particular issues with the advice of the Faculty of Child and Adolescent Psychiatry. The College felt there were advantages and disadvantages in a new measure. If a new measure were to be developed there would be “loss of ability to compare data”. However, lack of comparability needed to be balanced against the usefulness of a new instrument. The College stated that a variety of parameters should be incorporated into a measurement tool to reflect age, developmental, emotional, social, educational, physical and cultural differences. The College recommended a comprehensive measure but understood that comprehensiveness should be traded against brevity. The risks identified by the College concerned the misapplication or misinterpretation of results. The design of a sound instrument and the involvement of clinicians throughout development and implementation could minimise these risks.

A Dissenting View

Despite various qualifications and concerns, the great majority of stakeholders regard the use of some form of measurement system as important to child and adolescent mental health. Nevertheless, a number of stakeholders oppose the implementation of a measurement system. While these respondents were in a minority, their opinions amplify themes that emerged in the qualitative data provided by those who support outcome measurement. Dissenters are not convinced that a baseline-followup and concurrent feedback measurement system is an appropriate method of improving mental health services. While the minority do not necessarily oppose the ultimate application of some form of measurement, they consider there has been insufficient discussion concerning a number of key issues, a debate that should occur before measurement systems are designed, piloted or implemented. They contend that outcome measurement is primitive, and far from being able to provide what we really need to know. Child and adolescent mental health services cannot embark on outcome measurement with any confidence that such a system has integrity, or with any certainty as to "how (it) should be used, by whom and under what circumstances". Dissenters question whether a single outcome measurement system could address the heterogeneity of consumers, treatments and services. Standardised measurement will generate only undifferentiated data. Arguing against a "grand system", dissenters favour a "minimalist" approach that tailors outcome measurement to individual cases and targets only one or two domains.

The same dissenters oppose standardised outcome measurement because they believe such measurement would be unable to address the "qualitative client narrative" and would "fail to capture the full picture". As a result, "We could miss some of the human qualities that go with the distress and tragedies of mental health illness". Concerns were expressed that client and carer perceptions and their problems will be ignored or obscured by the incapacity of standard measures to capture the subtleties of individual cases. They are also concerned that a concurrent and baseline-followup system will be unfeasible due to the current limited infrastructure of child and adolescent mental health services. These stakeholders also believe that "the stupendous effort that would be required to impose some form of standard measure throughout Australia would not be justified by an improvement in the mental health of Australian children and young people". Concerns were expressed that a standardised measurement system would be outmoded after two years. Dissenters recommend further debate. At most, they suggest existing measures could be tentatively used to educate clinicians and promote constructive debate.

Barriers to Implementation

Clinicians

Clinician resistance was considered by a majority of stakeholders to be a serious potential barrier to the implementation of an outcome measurement system. Clinician resistance is thought to be generated by the following factors: anti-empiricism ("If we're just being told what to do, particularly if it's a bunch of academics who don't know what the real clinical world's like, there will be marked resistance"); resistance to, and fear of change; "paranoia"; concerns that outcome data will be used for “bench-marking” and “enterprise bargaining”; a detraction from direct service provision as a result of the burden of the system; a lack of expertise engendering defensiveness and repudiation; and if the system is poorly designed, the lack of relevance to clinicians of the concurrent and outcome data yielded.

Clients and Parents

Most stakeholders believe that, provided the system is not too burdensome, client -and parent-resistance will not be a problem. A number of respondents (including some consumers and their parents) believe consumers and their parents will welcome a measurement system. However, if parents or consumers feel stigmatised because they are accessing a mental health service, they may be resistant to measurement. Resistance is also thought likely if consumers and parents perceive the system as having the potential to reduce funding and service availability. Resistance is anticipated from special client groups such as homeless children.

Managers

Managers in general are not regarded as resistant. Most respondents believe that any resistance would relate to funding. Resistance is anticipated if managers perceive that outcome will become the sole basis for continued funding. Manager resistance is also anticipated if the system makes demands on existing budgets, especially if it is costly and cumbersome, or if managers must wait a long time to obtain funds to implement the system.

Funding and Resources

All stakeholder groups regard inadequate funding as a serious barrier to implementation. Many respondents predict that it will be expensive to develop, install, and sustain an outcome measurement system. There are concerns that inadequate funding will undermine successful implementation, particularly if the staffing of the system is inadequate. (“Mental health services would be unlikely to support outcome measurement systems that put an additional burden on existing staffing and finances.”)

Community Attitudes to Mental Health

Psychologists, nurses and parents believe that stigmatisation by the community could be a barrier to implementation.

Barriers Between States and Regions

The lack of uniformity and coordination, the “tribal warfare” between the states, and the autonomy of state and regional services, were seen as barriers to the implementation of a “national” or single system of outcome measurement. The difficulty of introducing a system that would fit comfortably with diverse systems in child health was also seen as a potential barrier.

Overcoming the Barriers to Implementation

Most stakeholders agreed that barriers to implementation would be overcome if stakeholders perceived the following to be the case:

- Stakeholders will contribute to the design and implementation of the system. (“If we’re simply just slaves of it, while three guys in Canberra and Brisbane can have a career on this, there’d be less willingness; but if there’s a collaborative effort, then people like me are going to be a whole hell of a lot more motivated and interested than if we have to be slaves to someone else’s enterprise.”)

- Stakeholders will understand and agree with how the data will be used.

- The system will be valuable to clinicians and their clients. (“Clinicians will make little investment and commitment if no value is seen to be coming from it.”)

- The data provided will be accessible, timely, and relevant to clinical practice. (“If clinicians can get data out of it, on a day-to-day basis, if they have got a question that is relevant to them, then they might be more committed.”)

- The measurement system will be well integrated into clinical practice.

- Adequate resourcing will be available to support installation and operation, in the form of money, equipment, dedicated staffing, and administrative support.

Stakeholders believe that the overcoming of resistance to a measurement system will require those who seek to implement it to apply the principles of change management. Respondents advise that the following are essential components of a strategy for change: stakeholder involvement, elegance of design, dedicated staffing, computerisation, training, demonstration projects, and the destigmatisation of mental health.

Stakeholder Involvement: “Involve the people. Involve them, involve them, and involve them! It’s the only way to getting around that (resistance).” “I just think there will be a lot of resistance on the part of clinicians and on the part of researchers if they’re not truly part of the effort. If they’re only the drones they’ll sabotage it. I certainly will. I won’t do it consciously, but I mean, this thing won’t work unless people get something for their labours.” Stakeholders advise that resistance will be mitigated if all stakeholders are involved in preliminary planning, design, and implementation. Open and honest communication is required through focus groups, staff briefings and briefings for clients and parents.

Design: A number of respondents consider that the way the system is designed will be crucial to its use. A measurement system should be elegant in presentation and simple to use. Questionnaires should be easy to complete. The clinical utility of the system should be transparent. The system should provide good feedback with data summaries and recommendations for treatment. The system should be an unintrusive part of routine practice. Consumers say that resistance would be reduced if they had a choice of assessment procedures. Consumers also see regular feedback to parents as important if parents have been involved in the young person’s treatment.

Dedicated Staffing: In the words of a clinician, “We can’t expect people to do more than they’re doing”. A dedicated staff member should have responsibility for the management of data.

Computerisation: Computerisation combined with software that presents data clearly and attractively will enhance clinician and client cooperation.

Training: Education and training are required for staff and managers.

National Demonstration Projects: Respondents suggest that national demonstration projects drawing on overseas experience will counter resistance.

Such projects could demonstrate the use and relevance of measurement systems.

Destigmatisation: A campaign of destigmatisation of mental illness within the community would mitigate barriers to implementation.

Computerisation

The vast majority of respondents see computerisation as extremely important, if not essential. Some respondents consider that computerisation is important not only in relation to concurrent and baseline-followup measurement, but also throughout child and adolescent mental health services. Respondents agree that computerised data collection should be part of routine practice. A computer should be installed in each clinician's office and incorporated into clinical work. Data collection, analysis, and feedback will thus be streamlined.

Stakeholders differ in their opinions regarding whether consumers and carers would be willing to engage in direct data entry (e.g., using touch-pad screens). Some expect positive responses. Others anticipate that consumers and carers will be resistant. Carers are less supportive of computerised data collection than of other forms of data collection.

Opinion is divided as to the role of clinicians in the collection of data. Some respondents think that data collection should be undertaken by the clinician in conjunction with the client. Others envisage data being provided by client and parent, independently of the clinician. Some respondents argue that administrative support should be provided. Others contend that clinicians should learn to manage a computerised system. On the other hand, some are concerned that computerisation will usurp the role of the clinician (e.g., by taking an intake history and conducting a mental state examination).

Education and Training

The majority of stakeholders regard education and training as crucial. ("Clinicians and managers can't make use of the system without training.") ("Everyone should realise what the data is and what it means.") ("They need to have a good understanding of all the statistical implications, and instrument reliability and validity.") A minority of stakeholders questions whether training is necessary. One respondent contended, ("Outcome data needs to be customised to meet the needs of clinicians who are working responsibly, not clinicians trained to respond to data.") Consumers believe that they should be educated to the extent that they have a clear understanding of the system.

Types of Training

Stakeholders should be trained to understand concurrent and outcome measurement and to operate the system. ("We need to tell them- clinicians - how to use it. A lot of them don't really understand things as simple as the CBCL. We need to teach them how to incorporate components of the measurement system in their interviews.") Training should address computing skills and the maintenance and use of data systems and information. Training should take place in clinical settings. The majority of respondents believe that initial training, on the job, is required. Continued training is seen as essential. ("Training should be done on a recurrent basis rather than a 'gee-whiz', come-in-and-tell-us-this-is-how-you-do-it-one-off thing. It has to be a continual process.")

Secondly, the deficiency of undergraduate training was a recurrent theme. There is little professional training in Australia in the use of outcome data or the design of concurrent and baseline-followup measurement. Some respondents consider that an understanding of outcome measurement systems should be part of undergraduate and postgraduate training. Others believe that the amount of training required depends on discipline. Contemporary social work training is said by some to encourage a "vehement opposition" to the use of empirical data. Social work students were described as a group requiring particular attention.

A third recommended approach to education and training involved modelling in demonstration centres where those who have experience in measurement systems provide training and guidance for others.

Stakeholder Perspectives on Risks and Ethical Issues

Overall, stakeholders do not anticipate that a measurement system poses significant risks or ethical concerns. The potential problems most often raised were the impact of a measurement system on confidentiality and the intrusiveness of a measurement system upon the therapeutic alliance between clients and clinicians.

Confidentiality

A significant number of senior mental health academics and professionals, clients and their parents are concerned about the potential loss of confidentiality involved in data storage, accessibility and data usage. Young people and parents are particularly concerned about such intrusions on their privacy.

Negative Impact on the Therapeutic Alliance

Of those stakeholders who perceive risks, a significant number, particularly those responding on behalf of Queensland Mental Health Services and senior professionals in the field, have concerns that a measurement system might interfere with the therapeutic alliance between client and clinician, especially if the client is severely disturbed.

The Intrusiveness of the Measurement System

A small number of stakeholders from several stakeholder groups expressed concern that precious clinical time could be lost to data collection and review. There were concerns that clinicians could become primarily “administrators” and “pen-pushers”. Some clinician stakeholders believe this would reduce staff morale and that “auditing madness could drive people out of the public sector because they’re too busy filling out forms”.

The Inadequacy of the Measurement System

Some stakeholders working in clinical settings fear the measurement system might be too unsophisticated to capture the complexity of child and adolescent mental health problems. “Treatment directives” promulgated by management would fail to address a problem of this magnitude. Stakeholders are also concerned that simplistic measurement will encourage a quasi-understanding of treatment and outcome, particularly by bureaucrats and policy-makers. This could foster administrative demands for clinicians to achieve an unrealistic outcome or to produce a good outcome in an unrealistic time.

Overvaluing the “Good Outcome”

Some stakeholders are concerned that services might focus excessively on outcome, as opposed to treatment processes. Services could become gripped by the need to achieve “good outcomes” to the detriment or exclusion of the client who has a severe and complex problem and who is unlikely to achieve a "good outcome" in the short term. Some stakeholders perceive that outcome data could be used to identify "successful" treatment modalities that would then be imposed on clinicians to the exclusion of other modes of treatment, thus infringing on clinician autonomy.

Risks to Funding and Resource Allocation

Some senior mental health academics and professionals are concerned that the implementation of a measurement system could consume already-limited resources. Stakeholders are concerned that government agencies and insurance companies might misinterpret data from a measurement system, failing to understand the complexity of mental health treatment and the difficulty of demonstrating a treatment effect, particularly in the short term. Misinterpretation could result in ill-informed or politically driven decisions about resource allocation, penalising and further marginalizing child and adolescent mental health services.

Medico-Legal Risks

A small number of stakeholders are concerned that measurement data could be used in cases alleging clinical negligence. Clinicians and services could become vulnerable to legal action, particularly if clinicians are alleged to have misused or failed to act upon feedback data.

Measurement Systems Currently Operating in Australia

All professional stakeholders consulted were asked whether they used outcome measurement in clinical settings, whether they knew of any measurement system that should be used in child and adolescent mental health, and whether they would adopt or encourage the adoption of a measurement system if one were to be developed. From responses to these questions, we identified services with a measurement system in place. We excluded outcome measurement systems associated with short-term research projects.

Eleven respondents, from four states, indicated that they routinely use outcome measurement. They were invited to provide information about their services' measurement systems. Two of the eleven respondents declined to be interviewed; one stated that his system was undergoing significant modification and he was not in a position to talk about it; while another cited time constraints as the main reason for declining to participate. One respondent could not be contacted. Six respondents were interviewed via telephone, and two returned a mailed response to the research team. Of the eight respondents who provided information, four were from New South Wales and three from Victoria and one from Queensland. All participants agreed that they and their location could be identified in the report. Each system will now be described.

Centre for Developmental Psychiatry, Monash University, Clayton, Victoria. Respondent : Professor Bruce Tonge

The Centre for Developmental Psychiatry at Monash University, Clayton, Victoria, uses outcome measurement designed to suit the treatment programs run by the centre. Although tailored for each program, the measurement systems share common features. Outcome measurement is based on standardised parent, self-report and professional checklists and interviews. All patients, parents or caregivers are included in the measurement of outcomes. While the different treatment programs vary in the time points at which they collect data, all do so at intake, at termination, and at least at six and twelve-month follow-up. At least one of the programs treating depressed young people collects data concurrently with treatment.

Professor Tonge indicated that measurement covers a range of domains including *diagnostic data*, *general symptomatology*, *adaptive functioning*, *goals of treatment*, *child and family environment*, *burden of care*, *quality of life*, *consumer satisfaction*, *treatment modality*. The specific measures used by four different programs were made available to us, showing that standard measures are used but that these vary across programs.

Professor Tonge reported that data is fed back to parents and sometimes children, to the funding source (e.g., NHMRC or the State Mental Health Branch), and to professionals (through conference presentations or publications). Outcome data inform the clinical management of treatment programs and are used to train clinicians and determine service profile and funding.

The Arndell Evaluative Research Program, Arndell Children's Unit, Royal North Shore Hospital, NSW. Respondent : Henry Luiker

According to the respondent, the Arndell Evaluative Program uses standardised outcome indicators collected from parents, teachers and children/adolescents. The computerised data management system is administered by non-specialist staff. The system evaluates day and residential care programs and is used for research, clinical assessment, and management.

Four standard instruments are used : the Child Behaviour Checklist, the Youth Self-Report, the Teacher's Report Form, and the Family Assessment Device. Domains covered include *diagnostic data*, *general symptoms*, *coping skills*, *functional competency, child and family environment*, *family resources*, and *burden of care*. Data are collected at intake, every six months during treatment, and six and twelve months after discharge by research personnel assisted by administration staff, using a specially designed computerised data management system.

Within a week of collection, outcome data are fed back to clinicians and teachers who work within the unit, using computer-generated graphs. Intake data are made available for use by the clinician in individual treatment planning. Six-monthly feedback data allow the clinician to get a sense of whether the child has improved six or twelve months after intake. At the service level, outcome data are used to determine the long-term effectiveness of the unit and to identify predictors of clinical outcome from the social demographic, and family, clinical, and institutional variables collected.

The service is planning to introduce measurement of treatment process factors such as *the quality of the interaction between treatment providers and clients* in order to examine the effect of treatment process on client outcome.

Rivendell Child, Adolescent and Family Services, Thomas Walker Hospital, Concord West, NSW. Respondent: Professor Joseph Rey

The Rivendell Child, Adolescent and Family Services uses a system that combines a continuous service audit with a patient satisfaction review. In use since 1988, the system is part of the Unit's overall quality assurance program. A list of all cases admitted in the previous three months is compiled and clinicians asked to indicate the current status of each case. The case is either closed, or admitted to the unit's residential or day programs, or treated in outpatient services. If the case is closed, the clinician rates the overall outcome as satisfactory, unsatisfactory or uncertain. Closed cases rated as having an unsatisfactory outcome are presented at a monthly peer review meeting. Staff discuss whether the case could have been handled differently and a better outcome achieved. Suggestions for improving future outcomes are discussed and circulated to staff (Plapp & Rey, 1994).

Following the clinician's notification that the case is closed, the parent completes a *parent satisfaction* questionnaire (PSQ). Cases for whom parents indicate low satisfaction are discussed at the monthly service audit meeting and an appropriate response decided upon. Once every six months, *parental satisfaction* responses are discussed at the peer review meeting, with the aim of improving services (Plapp & Rey, 1994).

The unit also uses routinely a *client satisfaction* questionnaire completed by adolescent in-patients. The questionnaire is mailed to the patient after discharge and returned questionnaires are reviewed regularly in quality improvement meetings. Information about the reliability, validity and usefulness of this questionnaire is not yet available.

Both *parental satisfaction* and *outcome* are routinely measured. The latter is assessed by means of a global rating ('satisfactory', 'unsatisfactory', 'uncertain') by the treating clinician based on whether the goals of assessment and/or treatment have been substantially achieved. Interrater reliability of these judgments has been shown to be moderate (kappa = 0.58) (Rey, Plapp & Simpson, in press). The *parent satisfaction* instrument used until recently was an abbreviated version of a widely used questionnaire (Nguyen, Attkisson & Segner, 1983). There was a significant but small agreement (27% better than chance) between clinicians' rating of outcome and parental satisfaction (Rey, Plapp & Simpson, in press). However, this instrument has been dropped following the analysis of data collected over a four-year period, which showed that the information gained through that questionnaire was too general and not particularly useful. As a result of that review, the Unit has designed a different questionnaire in which parents are asked whether they are satisfied or dissatisfied and to explain the reasons by choosing items from a list. The usefulness of the new approach will be evaluated in the near future.

Child and Adolescent Psychiatry Program, South East Sydney Area Health Service, NSW. Respondent: Associate Professor Stewart Einfeld

Professor Einfeld was interviewed by telephone. He described this system as a structured case presentation process that evaluates patient status before, after, and during treatment. The system has been in use since 1995. Conventional clinical assessment is undertaken alongside the outcome evaluation process.

Nine cases are audited every fortnight at three case review meetings. Cases are selected either by random audit or because they have had an unfavourable outcome. When a case comes up for audit, outcome data are collected using specific measures and assessments (e.g., the Achenbach checklists, the Global Family Environment Scale, the Global Assessment of Functioning, DSM IV Axes 1-5, a structured case presentation format, consumer satisfaction, and structured questionnaire items designed by the service). These measures and assessments cover the following domains: diagnostic data, general symptoms, adaptive functioning, goals of treatment, goal achievement, consumer satisfaction, social support, stressful events, treatment modality and therapeutic process.

Data are collected once only, except for cases with unfavourable outcomes. Such cases are reviewed again after three months. The system is used by both clinicians and management to highlight case issues that need further attention, to enhance treatment, and to identify educational needs or the need for system development.

The Alfred Child and Adolescent Mental Health Service, Melbourne, Victoria Respondent: Dr Allan Mawdsley

The Alfred Child and Adolescent Mental Health Service assesses client outcomes in three domains: *improvement in the severity of presenting complaints; goal achievement;* and *consumer satisfaction*.

At intake and case closure, clients and clinicians arrive at a consensus rating on a Likert scale (ranging from 1-9) of the severity of the three predominant presenting problems. The criterion for satisfactory improvement at case closure is an average of three or fewer scale points. At termination, both clients and clinicians are asked to rate the achievement of each goal along a scale of -10 to +10. Goal ratings are defined at assessment by client and clinician consensus as to what behaviours represent “best anticipated outcome” (+10), “satisfactory outcome” (+5), “current situation” (0), or “most unfavourable outcome” (-10). The criterion for excellent goal achievement is at least 50% average improvement for three goals. An assessment of goal achievement is also carried out every six months during treatment. Goals may be revised at this point. At the end of treatment, clients are asked to state in percentage terms how happy they have been with the service. A rating of at least 60% is required for client satisfaction to be considered satisfactory.

The aim of the system is to encourage staff to be customer-oriented, and to increase clinician effectiveness. Feedback from the outcome assessments is used for quality improvement on an individual case basis. At the service level, information is used to assess quality assurance targets.

Department of Child and Adolescent Psychiatry, Royal North Shore Hospital, NSW. Respondent: Dr Nick Kowalenko

This is a pre-and post-treatment system focussing on anxiety and mood disorders. Initial data are collected concerning the following domains : *diagnostic data, general symptoms, child and family functioning, goals of treatment, self-esteem, consumer satisfaction, interaction with other systems* and *treatment modality*. Data covering similar domains are collected at 6 months and at 12-months follow-up. Specific instruments are as follows: Child Behaviour Checklists, Youth Self-report (over 11 years), Structured Diagnostic Assessment of Anxiety Disorder for Children, Family Assessment Device, Fear Survey Schedule, Revised Manifest Anxiety scale, and a consumer satisfaction measure devised by the service. Data are collected from children aged from six to seventeen years and from parents/caregivers. All children and adolescents who are assessed complete the initial questionnaires. Those attending treatment programs complete the post-treatment and one-year follow-up questionnaires. Service director, clinic coordinator and clinicians within the service receive feedback data. Feedback is used in the following ways: to evaluate the effectiveness of structured multiple family group therapy; to modify the therapeutic program; and to evaluate the impact of treatment.

Maroondah Child and Adolescent Mental Health Service, Maroondah Hospital, Victoria. Respondent: Mr. Peter Brann

Since 1997, the Child and Adolescent Mental Health Service of Maroondah Hospital has been testing the Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA) as an outcome measurement instrument within a continuing quality improvement framework. This trial is being conducted in three community clinics, an adolescent residential unit, a mobile outreach adolescent service, and an adolescent day program.

HoNOSCA consists of 13 subscales and 2 optional items rated by the clinician on a scale of 0 to 4 (ranging from “no problem” to “very severe problem”). Clinicians complete the HoNOSCA form at first assessment and again at three months, six months, every six months after that, and at discharge. HoNOSCA subscales cover the following domains: *disruptive, antisocial or aggressive behaviour; overactivity, attention or concentration problems; non-accidental self-injury; alcohol and substance abuse; scholastic or language skills; physical illness or disability problems; hallucinations and delusions; non-organic somatic symptoms; emotional and related symptoms; peer relationships; self care and independence; family life and relationships;* and *school attendance*. Optional items include: *adequacy of information to carers about nature of the child’s difficulties*, and *lack of information about services or management of the child’s difficulties*).

Until recently, the trial has attempted to ascertain the psychometric properties of HoNOSCA, such as its validity and reliability. The trial has also focussed on the sustainability of routine outcome measurement. Clinicians are now provided with routine feedback information, by means of bar graphs showing subscale and total scores for the current and previous administrations of HoNOSCA. No information is yet available on how clinicians use feedback data. However, it is proposed that feedback data will inform clinicians concerning client progress and assist them to decide which cases require further attention (e.g., by peer review or re-assessment of treatment plans). Managers receive feedback in aggregate form and plan to use it to highlight strengths and weaknesses in service delivery, and to plan service modifications.

While the experience has been positive, the process of measuring outcome is in its infancy. Planned changes to the service’s outcome measurement include dropping the three-month data point and introducing a new measure, the

Strength and Difficulties Questionnaire (Goodman, 1997), which explores the perspectives of both parent and adolescent concerning the client's functional impairment and functional strengths.

School of Applied Psychology, Griffith University, Queensland
Respondent: Professor Mark Dadds

Outcome measurement is tailored to the various treatment programs. No specific description of the system was provided, but Professor Dadds indicated that outcome information is used for research and clinical work. Domains covered in measurement include *diagnostic data*, *general symptoms*, *child and family environment* and *consumer satisfaction*. A range of specific instruments is used according to age and symptom profile. Outcome data are collected from multiple sources including children, youth, parent or caregiver, teacher and mental health clinician, at intake and at termination of treatment. Feedback received by clinicians and the Clinic Director is used to monitor quality and evaluate service effectiveness. The measurement system aims to include all clients and their carers in outcome measurement.

It is evident that some services are committed to outcome measurement. However, systematic approaches to outcome measurement in child and adolescent mental health in Australia are few. Activities vary between agencies and services, and existing measurement systems appear to operate in isolation from each other.

CHAPTER IV

A CRITICAL REVIEW OF EXISTING INSTRUMENTS: METHODOLOGY AND RESULTS

Methodology

Information for the review of child and adolescent mental health measures was obtained via a comprehensive search using two literature databases (PsycInfo and MedLine). In addition, other sources including books and assessment resource catalogues were searched. The first step in locating relevant literature entailed searching keywords focusing on outcome and process in child and adolescent mental health. The second step involved narrowing the focus to specific domains of outcome and process measurement. Searches were limited to articles, books, and book chapters published in 1990 or thereafter, and to abstracts that included the root words “child*” (including “children” and “children’s”) or “adolesc*” (including adolescent, adolescents, and adolescence). The following keywords were combined with the keywords, “mental health”:

- outcome*
- measure*
- instrument
- assessment
- scale
- inventory
- schedule
- test
- rating
- checklist
- observation
- interview

The following keywords comprised the specific measurement domains:

- symptoms
- behaviour (and measure*; and outcome*)
- distress
- functioning
- resilience
- strength*
- competence or competencies
- parent-child relationship
- family functioning
- burden or caregiver strain
- stressful events
- coping skills

continued...

- self-esteem
- safety
- motivation to change
- satisfaction
- quality of life
- goals or goal attainment
- maltreatment

The initial search located approximately 500 abstracts. While sifting through these abstracts, we concurrently enlisted the aid of *The Directory of Unpublished Experimental Mental Measures* (Vols. 6 & 7) (Goldman & Mitchell, 1995, 1997), *The Consumer's Guide to Tests in Print* (Hammill, Brown & Bryant, 1992), seven assessment resource catalogues from publishers of mental measurements (see bibliography for references), and several review articles (especially Cross & McDonald, 1995). We prioritized the use of the secondary sources because they are comprehensive and contain rich reliability and validity data and source references (albeit the true source of instruments was often ambiguous). Instrument names with source references published before 1990 were listed and searched in the event that the instrument had been used in the 1990s. Any instrument that did not appear in an abstract after 1989 was excluded from this review, while those articles with instruments that were developed or used in the 1990s were retrieved. When appropriate, bibliographic citations were identified and retrieved.

We cannot claim that the search for every measure was exhaustive. Instrument names appear in the method sections of articles, but not always in the abstracts, and thus would be missed in an automated search process. Furthermore, psychometric data may be found in test manuals that, due to time and cost, we were unable to obtain. A total of 339 instruments were considered for inclusion and 188 instruments met criteria for review. The alphabetic list of excluded instruments and the main reason for the exclusion can be found in Appendix 5.

Exclusion Criteria

Several exclusion criteria were adopted primarily because their use is not feasible in typical clinic settings. Instruments were excluded from the review if they had any of the following characteristics:

- not developed or not found in the literature between 1990 and 1998

- used only with adults (aged 19 years and over)
- used only with children under 5 years of age
- requiring more than 30 minutes to administer
- requiring a special degree to administer
- solely teacher report, peer report, or requiring direct observation
- target population from a restricted age range (e.g., children between 6 and 9 years of age)
- designed solely for children with developmental delay or brain injury
- diagnosis- and problem-specific
- personality inventories
- projective tests

Domains

Two of the authors (LJAS & MSK), both with master's degrees in psychology, reviewed the measures using available studies. The measures were first placed in one of three domains. The first domain, *background* or *external factors*, are believed to *moderate* the effectiveness of treatment. These data provide information to the clinician that could be helpful in planning treatment. (e.g., *abuse history* and *family resources*). The second domain, *process variables*, are factors that are thought to *mediate* the effects of treatment. It is through these processes that treatment affects outcome. Since there has been little research in this area, there are few instruments. Examples of these areas are *therapeutic relationship* and *coping skills*. The third domain is outcome, the factors that clinician, family, client, and society look to as evidence of treatment effectiveness. Examples of outcomes include *alleviation of symptoms* and *enhancement of functional competence*. In some cases, a process or background domain can become an outcome of treatment, depending on the goals of treatment.

Moderating Domains

Family resources. Family resources data provide information about factors that affect the success of treatment. These measures assess the adequacy of different resources in the household (e.g., food, clothing, money, and time). Inadequate resources affect personal well being, family functioning, and parental commitment and could adversely affect parental involvement in child and family psychotherapy.

Stressful events. As with family resources, stressful events can affect a child's treatment progress. This domain consists of measures that identify external sources of stress ranging from major life events (e.g., death of a family member) to minor stressors (e.g., schoolwork). Stressful events could be assessed at baseline and concurrently to assist the clinician in identifying the past and current life difficulties that might contribute to mental health problems.

Safety of the living environment. A safe living environment engenders a sense of security. Security enhances mental health. Environments with safety concerns, such as exposure to violence, whether in the home or the community, place the child at risk for emotional and behavioral problems (Schwab-Stone, et. al., 1995) and can adversely affect treatment.

Maltreatment. Maltreatment measures assess current or past maltreatment of children by parents or other adults. Maltreatment includes physical, emotional, and sexual abuse and child neglect. These measures consist of parental report measures, child self-report measures, interviewer-based measures, and clinician ratings. A substantial proportion of maltreated youth exhibit internalizing and externalizing symptomatology (Wolfe, Gentile, & Wolfe, 1989). Maltreatment can affect the course of treatment. In a recent American survey, clinicians said that maltreatment was the domain about which they most wanted to receive information (Bickman, Rosof et al., in press).

Social support. Social support is a moderating domain; however, if treatment focuses on increasing social support, it can be considered a mediating domain. Social support measures assess the experience of intimacy, acceptance, companionship, and tenderness from peers, family members, and other adults. Research has shown that social support buffers children from stress (Wasserstein & La Greca, 1996). Thus, social support is an important factor in preventing child maladjustment, and a potential target for treatment planning.

Mediating Domains

Coping skills. Improved coping skills should lead to improvement in outcome. Coping skills instruments are process measures. Coping skills instruments measure coping patterns or styles, the frequency of coping behaviors, and the valence of coping (positive or negative).

Therapeutic processes. These measures assess different aspects of the therapeutic process: *therapeutic alliance, the participation of client and family in treatment*, *the motivation of the client to change*, *the therapy provided* and *the goals of treatment*. These processes are potential mediators of intermediate and distal outcome. For example, a good therapeutic alliance will increase attendance, decrease premature termination (Tryon & Kane, 1993) and improve client and parent satisfaction (Taylor, Adelman, & Kaiser-Boyd, 1986). Therapeutic process measures may also be linked to long term outcomes (e.g., *decrease in symptomatology*), either directly or indirectly through the aforementioned intermediate outcomes. Therapeutic processes are assessed by means of self-reports, parent reports, and clinician ratings.

Outcome Domains

Symptoms. The assessment of *symptoms* is the cornerstone of a measurement system. The symptoms children exhibit cause families to seek mental health services. Symptoms indicate the severity and course of the illness. Symptoms are assessed by structured and semi-structured interviews and behaviour checklists (Heflinger, 1992). Behaviour checklists provide a self-report format and extensive lists of problem behaviours and emotions. From the therapist’s perspective, symptom reduction is often the major goal of treatment (Strupp & Hadley, 1977).

Functional impairment. The assessment of functional impairment concerns those actions that indicate an inability to deal with the stresses and events of everyday life. Functioning has been assessed by parent and child self-report, interviewer and provider ratings, and indicators such as contact with the police and expulsion from school.

Functional competence. The assessment of functional competence concerns an individual’s adaptive functioning. Adaptive functioning is defined as behaviour effective in meeting personal and social demands and expectations within the subject’s environment (Harrison, 1989; Summerfelt & Bickman, 1994). Adaptive functioning involves the exceeding of normative social role expectations to the degree that it can be regarded as not merely an absence of impairment but also a strength (Kazdin, 1993; Reilly, 1996). Recently, functional competence has received increased attention (Lehnhoff, 1993; Stroul, 1996). Functional competence is most often assessed through self-report of parents and children.

Family functioning. Family functioning relates to how family members perform necessary roles and practical tasks, adapt to problems, and communicate with one another, in such a way that they are able to live together (Fobair & Zabora, 1995). The relationship between family functioning and a child's emotional and behavioural problems is a transactional process, the family system both affecting, and being affected by, the behaviour of individual family members (Friesen & Koroloff, 1990). Several family functioning instruments have been developed for the evaluation of families in family therapy. None has been developed for outcome monitoring. Family functioning may also be a background or process domain.

Satisfaction. Satisfaction with services is either an outcome domain or a process domain (if satisfaction mediates the effectiveness of treatment). Satisfaction is associated with care-seeking behaviour and adherence to treatment (Ruggeri, 1994). Satisfaction is measured globally or according to satisfaction with specific aspects of treatment (e.g., adequacy of facilities, accessibility of treatment). Satisfaction measures tend to be self-report measures completed by parents or adolescents.

Self-esteem. Self-esteem measures assess a child's attitudes, feelings, and knowledge of his/her abilities, skills, appearance, and social acceptability (Byrne, 1983). These measures tend to be self-reports. Children with high self esteem are likely be better students, to have less anxiety and depression, and to display better physical health and social relationships (Gilberts, 1983).

Quality of life. Quality of life refers to a general feeling of subjective well-being and satisfaction with life. It covers multiple areas of life: physical, psychological, cognitive, affective, social, family, and economic (Ivan & Glazer, 1994). Life satisfaction is negatively correlated with psychopathology such as depression (Headey, Kelley, & Wearing, 1993).

Goal attainment: Goal attainment scaling (GAS) was developed by program evaluators to assess change in individual clients. At intake or baseline, the youth's status is estimated. The GAS evaluator then works alone or with the child and family to predict reasonable client-specific outcomes at termination and follow-up (Kiresuk & Lund, 1982). This technique is discussed at greater length in Chapter V (Goals).

Multidimensional. A number of the measures we found related to more than one domain. It would be advantageous to identify a single instrument that assesses and integrates several domains briefly and effectively. We found several instruments that collect information from different sources (parents, children, and clinicians) across several domains (e.g., *functional impairment*, *strengths*).

Evaluation of Measures

As the next step in the selection process, measures were examined with regard to their psychometric characteristics. All available literature was reviewed concerning each instrument. The raw data used in the evaluation of each instrument is summarized in Appendix 5.

Each of the relevant instruments was reviewed according to 29 criteria (summarized in Appendix 5 as table columns). The same categories were then weighted by the values shown in Table V.1, “Coding Key” (e.g., higher reliability coefficients are given more credit). It is important to note that the psychometric information provided is not immutable and “part” of the instrument but dependent on the context in which the instrument is used. One cannot assume that an instrument with a published high reliability coefficient is reliable under circumstances different from those in which the reliability was originally tested. Reviewers analyzed measures in accordance with the following criteria:

1. **Cultural sensitivity: Evidence of bias.** Measures were examined for cultural bias. Cultural bias means that the measure has different psychometric properties in different cultural groups. In almost all cases, cultural bias had been examined only with respect to American Hispanics or African-Americans.

2. **Cultural sensitivity: Cultural norms.** Measures were examined concerning whether they had norms for different cultural groups. Norms refer to the scores obtained from testing a representative sample of the cultural group with the instrument. If separate cultural norms exist, it might be possible to correct for cultural bias.

3. Cultural sensitivity: Used in Australia. Since the meaning of measures can vary with culture, we examined whether measures had been used in Australia. Mere use in Australia, however, does not necessarily mean that the instrument is culturally fair in Australia.

4. Suitability: Cost. Measures were examined for the cost of an introductory kit/manual and price per 100 forms. Cost is presented in U.S. dollars.

5. Suitability: Time. Measures were examined for time-to-administer in minutes (upper bound). For some measures, time-to-complete was not indicated in published information. To estimate completion time, we derived a formula from the selected measures for which we had time data : Time (in minutes) = 4.4 minutes + (.09) (number of items). This formula has a r2 of .81. The intercept represents the startup time required. Comprehensive measures may require multiplication of this startup time whenever they change response format or instructions (e.g., from a yes-no format to a Likert scale).

6. Suitability: Training. Measures were examined for the amount of training required to administer and score the measure.

7. Suitability: Degree/experience. Measures were examined for any requirements relative to training background or assessment experience required to administer them.

8. Suitability: Reading level. Measures were evaluated for their reported reading levels (i.e., the reading grade level required to complete the measure). Reported reading levels were often based on formulas that evaluate sentence length, word difficulty, and number of syllables per word.

9. Suitability: Computer software available. Measures were reviewed as to whether or not they had computer software available for administration or scoring or interpretation.

10. Reliability: Test-retest reliability tested within 2 weeks. We examined test-retest within two weeks because, as the time period increases, many more variables other than the reliability of the instrument can intervene to change responses. Thus, a measure may appear unreliable when in fact it is reasonable for the construct to change. Over a shorter period of time, it is

anticipated that a subject would respond consistently to the measure (less time for confounding variables).

11. Reliability : Test-retest coefficient. Measures were examined for whether subjects responded consistently to the measure over a period of time. The coefficient, typically a Pearson correlation, represents agreement between different administrations of the instrument to the same subject. Criteria for acceptable test-retest reliability ratings were based on Cross and Mcdonald (1995), Hammill, Brown, and Bryant (1992), and Rossi and Freeman (1993).

12. Reliability : Internal Consistency. Measures were examined concerning inter-item reliability coefficient (typically Cronbach's alpha) for entire scales and subscales. If the items of an instrument form a scale, there should be high inter-item correlations among the items. Criteria for the acceptability of internal consistency estimates were based on Nunnally (1978).

13. Reliability: Cross-informant agreement. Child and adolescent assessment is facilitated if information is gathered from multiple perspectives (e.g., the child/adolescent, parents, teachers, therapists, peers and other individuals who know the child well). Because of the contrasting views and specificity of contextual factors, multi-informant agreement can span a wide range. It was not possible to evaluate agreement for a wide variety of measures and informant dyads. Instead, we indicate whether any cross-informant data have been reported.

14. Reliability : Interrater agreement. The degree of agreement or consistency between ratings by similar raters (e.g., clinicians, parents, teachers) was examined. Criteria for acceptable Pearson, intraclass, and Kappa coefficients were based on Orwin (1994) and Nunnally (1978).

15. Content Validity : Theory-based items. Content validity involves the systematic examination of the test content to determine whether it covers a representative sample of the behaviour domain to be measured (Anastasi, 1988). We reviewed the literature for information on how measures were developed and determined whether items had been generated in accordance with theory.

16. Content validity : Judgments. We reviewed measures to determine whether the process of development involved experts who rated the relevance, representativeness or quality of items.

17. Content validity : Rules. Content validity was checked. Were rules or algorithms used to determine item relevance or to reduce the number of items, thus producing a content-valid and psychometrically sound scale? Item discrimination, factor analysis, classical item analysis, and the application of criteria concerning the acceptability of items were noted.

18. Content validity : Respondent feedback. We determined whether respondents (e.g., children, parents, teachers clinicians or peers) generated items or reviewed items for inclusion and exclusion.

19. Construct validity : Factor analysis. We examined whether factor analytic studies had been conducted and whether any of these studies disconfirmed the theorized structure of the measure.

20. Construct validity : Convergent validity. Convergent validity is the degree to which a measure is related to a similar measure or criterion that theoretically represents the same or a similar construct. We reviewed validity studies and studies using child and adolescent measures to evaluate the convergence of theoretically-linked criterion measures or scales. Both concurrent and predictive validity estimates were examined. In evaluating convergent validity, we rated the one criterion or measure that most highly correlated with the scale. The acceptability of association coefficients was based on Cross and McDonald (1995).

21. Construct validity : Divergent validity. Validity evidence can also be obtained by estimating the relationship between the measure in question and a criterion measure or scale with which it should not be associated. In evaluating divergent validity, we rated the one criterion or measure that had the lowest correlation with the scale and was hypothesized not to be related. The acceptability of divergent validity estimates was based on Cross and McDonald (1995).

22. Construct validity : Social desirability. Social desirability is the tendency of individuals to respond to a measure in a socially biased way. Certain measures and measurement contexts are susceptible to social desirability effects (e.g., a desire to look good or please the researchers by exaggerating,

minimizing, denying, or lying). If data were available, we evaluated the degree to which the measure correlated with socially desirable responding. Because this criterion is a sub-category of divergent validity, we again deferred to Cross and McDonald's (1995) acceptability ratings.

23. Construct validity : Group differences. We reviewed data regarding the measure's ability to distinguish between groups of individuals who should differ on that construct. Because data on the capacity of measures to differentiate different groups are scarce, we could only determine whether studies reported significant discriminations and the number of studies or group pairs compared.

24. Construct validity : Sensitivity / Specificity. Sensitivity is a measure's ability to identify individuals who truly manifest the problem or disorder of interest. An insensitive measure will have difficulty detecting the construct of interest. Specificity is a measure's ability to accurately distinguished individuals who do not manifest the problem or disorder of interest. A non-specific measure will lead to false positive classifications. We examined the literature for sensitivity/specificity data.

25. Norms : Current. The literature was reviewed to determine whether up-to-date normative data have been reported. Norms were considered current if collected within the last 20 years.

26. Norms : Number of normative samples. Multiple normative samples are important because they replicate earlier studies or add new information about normative scores for different groups. We reported the total number of normative samples in the literature.

27. Developmental sensitivity: Number of forms. Cognitive and social development affect the readability, comprehensibility, reliability, and meaning of measures. For instance, some measures are inappropriate for young children due to their cognitive immaturity or lack of comprehension; whereas others are inappropriate for older children/ adolescents because the format or items are too simple. The items may be suitable for children but not for adolescents (e.g., because they involve questions about sexual relations). The number of forms developed for specific age-groups is indicative of authors' attention to issues of developmental sensitivity.

28. **Developmental sensitivity: Age effects.** We reviewed data on measures to determine whether scores or reliability estimates differed by age or age group.

29. **Developmental sensitivity: Norms.** If age effects are present, data that specify normative scores across ages and age groups are particularly important. We reviewed normative data, if available, and determined whether separate norms were reported by age.

The next step in the evaluative process was to work with the data in Appendix 5 in order to produce summary evaluative tables for each instrument. While Appendix 5 provides the raw information needed to evaluate each instrument, it does not apply standards to the 29 criteria. In order to do so, the reviewers used the simple coding key shown in Table IV.1, below. The summary tables that follow use asterisks and other symbols to indicate the quality of the measure for the given characteristic. The key for describing the quality of the measures was created from multiple literature sources and from clinical and research experience. Test-retest reliability ratings were based on Cross and McDonald (1995), Hammill, Brown, and Bryant (1992), and Rossi and Freeman (1993). Internal consistency ratings were based on Nunnally (1978). Interrater agreement ratings were based on Orwin (1994) and Nunnally (1978). Convergent and divergent validity ratings were based on Cross and McDonald (1995).

Table IV-1 — Coding Key

Criterion	Coding Method	
Cultural sensitivity: Evidence of bias	Y	Evidence suggesting scale is statistically biased relative to cultural issues
	N	Evidence suggests lack of bias
Cultural sensitivity: Cultural norms	Y	Norms for different cultural groups
	N	No norms for different cultural groups
Cultural sensitivity: Used in Australia	Y	Scale used or tested in Australia
Suitability: Cost		Introductory kit/manual plus price per 100 forms – in US dollars
Suitability: Time		Time to administer in minutes (upper bound)
Suitability: Training	0	None
	t	some-2 hours
	tt	3-4 hours
	ttt	5-8 hours
	tttt	2-4 days
	ttttt	5 or more days
Suitability: Degree/experience requirements	0	None
	d	Minimum Bachelor's degree or equivalent
	dd	Minimum Master's degree or equivalent
	ddd	Minimum Doctoral degree or equivalent
	&	Measure requires general assessment experience or training background
Suitability: Reading level		Grade level
Suitability: Computer software available	Y	Software available
	N	Software not available
Reliability: Test-retest within 2 weeks	Y	Test-retest examined for a time period of 2 weeks or less
	N	Test-retest examined for time period > 2 weeks
	*	Test-retest examined for unspecified time period
Reliability: Acceptability of test-retest reliability for shortest test-retest period	*	Unacceptable reliability coefficient any subscale or studies < .60
	**	Barely acceptable reliability coefficient, all subscales or studies >.59
	***	Somewhat acceptable reliability coefficient, all subscales or studies >.70
	****	Mostly acceptable reliability coefficients, all subscales or studies >.80
	*****	Highly acceptable reliability coefficient, all subscales or studies >.89
		(Cross & McDonald, 1995; Hammill, Brown, & Bryant, 1992; Rossi & Freeman, 1993)

Table IV-1 — Coding Key cont'd

Criterion	Coding Method	
Reliability: Acceptability of internal consistency	*	Unacceptable reliability coefficient any subscale or studies < .70
	**	Barely acceptable reliability coefficient, all subscales or studies >.69
	***	Somewhat acceptable reliability coefficient, all subscales or studies >.80
	****	Mostly acceptable reliability coefficients, all subscales or studies >.90
	*****	Totally acceptable Reliability coefficient, all subscales or studies >.94
		(Nunnally, 1978)
Reliability: Cross-informant agreement	Y	Any cross-informant study conducted (parent-parent, parent-child, parent-clinician, child-clinician
	N	No cross-informant studies conducted
	●	See appendix for actual reliability coefficients
Reliability: Interrater agreement (similar informant type)	k	Indicates Kappa used for coefficient of agreement
	*	k < .40, ICC < .50, Pearson r < .70
	**	k = .40-.59, ICC = .50-.74, Pearson r = .70-.79
	***	k = .60-.74, ICC = .75-.89, Pearson r = .80-.89
	****	k > .74, ICC > .89, Pearson r > .89
		(Cooper & Hedges, 1994; Nunnally, 1978)
Content Validity: Theory-based items	Y	Items generated based on theoretical specification of content: includes use of literature
	N	Items not generated from theory
Content Validity: Expert judgments	Y	Experts rated relevance, representativeness, or quality of items
	N	Items not evaluated by experts
Content Validity: Rules to generate and exclude items	Y	Rules or algorithms used to generate or exclude items
	N	Rules or algorithms not used to generate or exclude items
Content Validity: Subject matter experts	Y	Subject matter experts generated or excluded items (children, parents, teachers, clinicians, etc.)
	N	Experts not used to generate or exclude items
Construct Validity: Factor analysis conducted	Y	Any factor analytic studies conducted with measure
	N	No factor analytic studies conducted
	-D	Disconfirmed within primary sample: items did not factor into hypothesised dimensions.
Construct Validity: Convergent validity		Used the one criterion or measure most highly correlated with the scale
	*	Unacceptable: any subscale correlated < |.50| in any study
	**	Barely acceptable: lowest correlation of any subscale or study > or = |.50|
	***	Highly acceptable: all subscales theorised to be related to the construct correlated > |.75|
	N	No convergent validity tests conducted
		(Cross & McDonald, 1995)

Table IV-1 — Coding Key cont'd

Criterion	Coding Method	
Construct Validity: Divergent validity		Used the one criterion or measure that had the lowest correlation with the scale and was hypothesised to not be related
	*	Unacceptable: highest correlation of any subscale or study with divergent construct > |.24|
	**	Barely acceptable: highest correlation of any subscale or study with divergent construct < |.25|
	***	Highly acceptable: highest correlation of any subscale or study with divergent construct < |.11|
	N	No divergent validity tests conducted (Cross & McDonald, 1995)
Construct Validity: Social desirability		Correlation with lie scale, validity scale, or social desirability measure
	*	Unacceptable: highest correlation of any subscale or study with SD > |.24|
	**	Barely acceptable: highest correlation of any subscale or study with SD < |.25|
	***	Highly acceptable: highest correlation of any subscale or study with SD < |.11|
	N	No social desirability tests conducted
Construct Validity: Group difference evidence	Y	Any study found that the measure significantly discriminates between known groups
	N	Measure not found to significantly discriminate between known groups
	(1,2…)	Number of known group pairs significantly differentiated by the measure
Construct Validity: Sensitivity/specificity	Y	Any studies conducted with measure to examine its sensitivity or specificity in classification
	N	No studies conducted to examine sensitivity or specificity
Norms: Current	Y	Any norms within last 20 years
	N	No current norms
Norms: Number of normative samples		Number of normative samples (add up ones from clinical sample, inpatient, etc.)
Developmental Sensitivity: Forms		Number of forms (i.e. different forms for different ages)
Developmental Sensitivity: Age effects	Y	Any statistical effects of age on scores
	N	Scores not found to be effected by age
Developmental Sensitivity: Norms	Y	Any age-specific norms
	N	No age-specific norms

Note: Y = yes; N = no; ICC = intraclass correlation coefficient.

Measure	Cultural Sensitivity: Evidence of bias	Cultural Sensitivity: Cultural norms	Cultural Sensitivity: Used in Australia	Suitability: Cost	Suitability: Time	Suitability: Training	Suitability: Degree / experience	Suitability: Reading level	Suitability: Computer software available	Reliability: Test-retest within 2 weeks	Reliability: Test-retest coefficient	Reliability: Internal consistency	Reliability: Cross-informant agreement	Reliability: Interrater agreement	Content Validity: Theory-based items	Content Validity: Judgments	Content Validity: Rules	Content Validity: Respondent feedback	Construct Validity: Factor analysis	Construct Validity: Convergent validity	Construct Validity: Divergent validity	Construct Validity: Social desirability	Construct Validity: Group differences	Construct Validity: Sensitivity / Specificity	Norms: Current	Norms: Number of normative samples	Developmental sensitivity: Forms	Developmental sensitivity: Age effects	Developmental sensitivity: Norms
Coping Skills Measures																													
Adolescent-Coping Orientation for Problem Experiences Scale	Y											*							Y-D	*									
Analysis of Coping Style					15											Y							Y1						
Coping Inventory for Adolescents												*			Y		Y		Y										
Coping Responses Inventory				179	15		&	6																					
Coping Strategies Inventory										Y	**	**			Y				Y	*			Y3						
Kidcope										Y	*				Y, but see app					*			Y2						
Ways of Coping Scale					See app							*																Y	
Family Functioning																													
Adolescent Family Life Satisfaction Index												***				Y			Y	**									
Caregiver Strain Questionnaire					10							**							Y	*									
Child and Adolescent Burden Assessment					10	t				Y	**	***			Y			Y	Y-**D**				Y1						
Children's Report of Parental Behavior Inventory	Y											*			Y		Y						Y3						
Cleminshaw-Guidubaldi Parent-Satisfaction Scale	N			206	30		dd					**							Y						Y			N	

Measure	Cultural Sensitivity: Evidence of bias	Cultural Sensitivity: Cultural norms	Cultural Sensitivity: Used in Australia	Suitability: Cost	Suitability: Time	Suitability: Training	Suitability: Degree / experience	Suitability: Reading level	Suitability: Computer software available	Reliability: Test-retest within 2 weeks	Reliability: Test-retest coefficient	Reliability: Internal consistency	Reliability: Cross-informant agreement	Reliability: Interrater agreement	Content Validity: Theory-based items	Content Validity: Judgments	Content Validity: Rules	Content Validity: Respondent feedback	Construct Validity: Factor analysis	Construct Validity: Convergent validity	Construct Validity: Divergent validity	Construct Validity: Social desirability	Construct Validity: Sensitivity / Specificity	Construct Validity: Group differences	Norms: Current	Norms: Number of normative samples	Developmental sensitivity: Forms	Developmental sensitivity: Age effects	Developmental sensitivity: Norms
Family Functioning *cont'd*																													
Clinical Rating Scale for the Circumplex Model of Marital and Family Systems														***															
Family Adaptability and Cohesion Scales II	Y									N	***	*	• Y	*		Y			Y	*			Y4		Y	2			
Family Adaptability and Cohesion Scales III	See app		Y		20					N	***	*	• Y		Y		Y		Y	***	*	*	Y4		Y	Many: see app			
Family APGAR	Y									Y	****	**			Y					***			Y1	Y	N				
Family Environment Scale	Y				10					N	**	*			Y		Y	Y	Y-**D**	**	*	*	Y2		Y	2			
Family Expressiveness Questionnaire											***								Y										
Family Functioning Index					15								• Y	*				Y: see app		***			Y1		N				
Family Functioning Questionnaire												* see app	• Y	**	Y		Y												
Family Functioning Scales												*																	
Family Functioning Style Scale					See app							****							Y	**									
Family Intervention Scale					3									*; see app				Y					Y1						
Family Life Questionnaire										*	**	***								*									
Family Satisfaction												***	• Y	*						**									

Measure	Cultural Sensitivity: Evidence of bias	Cultural Sensitivity: Cultural norms	Cultural Sensitivity: Used in Australia	Suitability: Cost	Suitability: Time	Suitability: Training	Suitability: Degree / experience	Suitability: Reading level	Suitability: Computer software available	Reliability: Test-retest within 2 weeks	Reliability: Test-retest coefficient	Reliability: Internal consistency	Reliability: Cross-informant agreement	Reliability: Interrater agreement	Content Validity: Theory-based items	Content Validity: Judgments	Content Validity: Rules	Content Validity: Respondent feedback	Construct Validity: Factor analysis	Construct Validity: Convergent validity	Construct Validity: Divergent validity	Construct Validity: Social desirability	Construct Validity: Group differences	Construct Validity: Sensitivity / Specificity	Norms: Current	Norms: Number of normative samples	Developmental sensitivity: Forms	Developmental sensitivity: Age effects	Developmental sensitivity: Norms
Family Functioning *cont'd*																													
Family Stressor Scale												*																	
Global Family Environment Scale	Y		Y			tt				*	*****			**						***			Y2						
Issues Checklist												*																	
McMaster Family Assessment Device	Y				20					Y	**	*							Y	*	*	**	Y2	Y	Y				
Modified Impact on Family Scale										Y	*****	****																	
O'Leary-Porter Scale												**																	
Parent Perception Inventory												**																	
Parent-Adolescent Attachment inventory												***																	
Parent-Adolescent Communication Inventory or Scale	Y											*	• Y	*						*									
Parental Acceptance-Rejection Questionnaire	Y									N	**	**							Y										
Parental Nurturance Scale										*	*****	****								*									
Parent-Child Scales										N	**	*											Y2						
Quality of Parent-Child Relationships scale												***																	
Self-Report Family Inventory										N	*	***			N	N	N		Y	**	*				N				
Structural Family Interaction Scale-Revised										N	****	**								**	*		Y2						
Youth-Parent Conflict Scale												***																	

Measure	Cultural Sensitivity: Evidence of bias	Cultural Sensitivity: Cultural norms	Cultural Sensitivity: Used in Australia	Suitability: Cost	Suitability: Time	Suitability: Training	Suitability: Degree / experience	Suitability: Reading level	Suitability: Computer software available	Reliability: Test-retest within 2 weeks	Reliability: Test-retest coefficient	Reliability: Internal consistency	Reliability: Cross-informant agreement	Reliability: Interrater agreement	Content Validity: Theory-based items	Content Validity: Judgments	Content Validity: Rules	Content Validity: Respondent feedback	Construct Validity: Factor analysis	Construct Validity: Convergent validity	Construct Validity: Divergent validity	Construct Validity: Social desirability	Construct Validity: Group differences	Construct Validity: Sensitivity / Specificity	Norms: Current	Norms: Number of normative samples	Developmental sensitivity: Forms	Developmental sensitivity: Age effects	Developmental sensitivity: Norms
Family Resources Measures																													
Family Needs Scale					See app							*****				See app			Y	*					N				
Family Needs Survey					See app					N	*		• Y						Y-**D**				Y1	Y	N				
Family Resource Scale										N	*	****			Y	Y			Y	**					N				
Support Functions Scale										N	**	***				See app			Y	*					N				
Functioning Competence Measures																													
Behavioral and Emotional Rating Scale		Y		142	10		&			*	****	***		***		Y	Y	Y	Y	*			Y1		Y	2			Y
CBCL Social Competence Scale	N; See app	N		143	10			5		Y	P**** A***	*	• Y	****					Y-**D**	**			Y1	Y	Y	P4; A4	Y	Y	Y
Child Behavior Rating Scale (Rochester Social Problem Solving Core Group, 1980)										*	***																		
Health Resources Inventory										N	***	**	• Y						Y	*									
Normative Adaptive Behavior Checklist					22							See app													Y; see app				

Measure	Cultural Sensitivity: Evidence of bias	Cultural Sensitivity: Cultural norms	Cultural Sensitivity: Used in Australia	Suitability: Cost	Suitability: Time	Suitability: Training	Suitability: Degree / experience	Suitability: Reading level	Suitability: Computer software available	Reliability: Test-retest within 2 weeks	Reliability: Test-retest coefficient	Reliability: Internal consistency	Reliability: Cross-informant agreement	Reliability: Interrater agreement	Content Validity: Theory-based items	Content Validity: Judgments	Content Validity: Rules	Content Validity: Respondent feedback	Construct Validity: Factor analysis	Construct Validity: Convergent validity	Construct Validity: Divergent validity	Construct Validity: Social desirability	Construct Validity: Group differences	Construct Validity: Sensitivity / Specificity	Norms: Current	Norms: Number of normative samples	Developmental sensitivity: Forms	Developmental sensitivity: Age effects	Developmental sensitivity: Norms
Functioning Impairment Measures																													
Child and Adolescent Functional Assessment Scale	Y				30; see app	see app	0			*	***	*		*						**			Y2				2	Y	
Children's Global Assessment Scale: Clinician and lay interviewer versions						See app				N	****		• Y	**; see app						***			Y4	Y				N	
Columbia Impairment Scale					5					*	P**** Ch***									P** Ch*			Y2						
Personal Problems Checklist-Adolescent (PPC-A) and Children's Problems Checklist (CPC)				74 (both)		0		PPC-A; 7																					
TOCA-R												****							Y						Y				
General Symptoms Measures																													
Achenbach scales CBCL YSR	N	N		143	17			5	Y	Y	P** A*	*	• Y	*	Y		Y		Y	***			Y5	Y	Y	4	2	Y	Y
Behavior Dimensions Rating Scale				239	10	0				*	****	****		*		Y			Y	See app			Y1	Y	Y	2		Y	Y
Behavioral Problems Index												***															2		
Behavior Rating Profile-2nd edition		Y		242	20					Y	*	**	N		Y			Y		N	Y				Y	3; see app		Y	
Bristol Social Adjustment Guides					15							*																	

Measure	Cultural Sensitivity: Evidence of bias	Cultural Sensitivity: Cultural norms	Cultural Sensitivity: Used in Australia	Suitability: Cost	Suitability: Time	Suitability: Training	Suitability: Degree / experience	Suitability: Reading level	Suitability: Computer software available	Reliability: Test-retest within 2 weeks	Reliability: Test-retest coefficient	Reliability: Internal consistency	Reliability: Cross-informant agreement	Reliability: Interrater agreement	Content Validity: Theory-based items	Content Validity: Judgments	Content Validity: Rules	Content Validity: Respondent feedback	Construct Validity: Factor analysis	Construct Validity: Convergent validity	Construct Validity: Divergent validity	Construct Validity: Social desirability	Construct Validity: Group differences	Construct Validity: Sensitivity / Specificity	Norms: Current	Norms: Number of normative samples	Developmental sensitivity: Forms	Developmental sensitivity: Age effects	Developmental sensitivity: Norms
General Symptoms Measures *cont'd*																													
Eyberg Child Behavior Inventory										*	****	*****													Y				
Ontario Child Health Study scales										N	*									*			Y6	Y					
Parent Daily Report										*	****	*	Y	***															
Pediatric Symptom Checklist	Y				5					Y	****	***		k**						**			Y7	Y; see app	Y				See app
Revised Behavior Problem Checklist	Y	N		257	20	&				N	*	**	Y	*						*	***		Y6	Y	Y	1			
Vermont System for Tracking Client Progress			N									**							Y-**D**	**			Y1						
Walker Problem Behavior Identification Checklist										*	*	*****						Y	Y						Y	1			Y
Goals Measures																													
Adolescent Goal Attainment Scale													• Y							*									
Family Goal Recording														K****	Y					**; see app									

Measure	Cultural Sensitivity: Evidence of bias	Cultural Sensitivity: Cultural norms	Cultural Sensitivity: Used in Australia	Suitability: Cost	Suitability: Time	Suitability: Training	Suitability: Degree / experience	Suitability: Reading level	Suitability: Computer software available	Reliability: Test-retest within 2 weeks	Reliability: Test-retest coefficient	Reliability: Internal consistency	Reliability: Cross-informant agreement	Reliability: Interrater agreement	Content Validity: Theory-based items	Content Validity: Judgments	Content Validity: Rules	Content Validity: Respondent feedback	Construct Validity: Factor analysis	Construct Validity: Convergent validity	Construct Validity: Divergent validity	Construct Validity: Social desirability	Construct Validity: Group differences	Construct Validity: Sensitivity / Specificity	Norms: Current	Norms: Number of normative samples	Developmental sensitivity: Forms	Developmental sensitivity: Age effects	Developmental sensitivity: Norms
Maltreatment Measures																													
Adolescent Abuse Inventory												*																	
Child Abuse Potential Inventory	Y				20			3		Y	*****	**; see app			Y		Y		Y–D	***			Y4	Y	Y			Y	
Child Report of Treatment Issue Resolution												****				Y			Y	**		*							
Sexual Abuse Fear Evaluation Subscale												**																	
Trauma Symptom Checklist				241	20		dd&		Y			**								***					Y	4			
Multidimensional Measures																													
Adolescent Treatment Outcomes Module					28			4		*	**	*			Y	Y				P** A*			Y4	Y					
Behavior Assessment System for Children	N			125	30; see app				Y	N	Tchr**** P** C***	Tchr*** P* C*	• Y	*	Y		Y	Y	Y	** ; see app			N		Y	2	3	Y	
Child Behavior Rating Scale										Y	***		Y	*															
The Health of the Nation Outcome Scales for Children and Adolescents: Version 6.0																				**									
Matson Evaluation of Social Skills with Youngsters										Y	*	***							Y	*			Y1		Y				
Mental Status Checklist				148	20	0						***	• Y						Y	*	*			Y			3		

Measure	Cultural Sensitivity: Evidence of bias	Cultural Sensitivity: Cultural norms	Cultural Sensitivity: Used in Australia	Suitability: Cost	Suitability: Time	Suitability: Training	Suitability: Degree / experience	Suitability: Reading level	Suitability: Computer software available	Reliability: Test-retest within 2 weeks	Reliability: Test-retest coefficient	Reliability: Internal consistency	Reliability: Cross-informant agreement	Reliability: Interrater agreement	Content Validity: Theory-based items	Content Validity: Judgments	Content Validity: Rules	Content Validity: Respondent feedback	Construct Validity: Factor analysis	Construct Validity: Convergent validity	Construct Validity: Divergent validity	Construct Validity: Social desirability	Construct Validity: Group differences	Construct Validity: Sensitivity / Specificity	Norms: Current	Norms: Number of normative samples	Developmental sensitivity: Forms	Developmental sensitivity: Age effects	Developmental sensitivity: Norms
Multidimensional Measures *cont'd*																													
Ohio Youth Problems, Functioning, and Satisfaction Scales		N	N		10	t						P** Y** CM*			Y			Y		P*** Y* CM**			Y4		Y	2			N
Paddington Complexity Scale																				**									
Questionnaire developed for specific study (Slee & Rigby, 1993)																			Y	*									
Social Adjustment Inventory for Children and Adolescents					30	tt	dd&						• Y	K****					Y	**	**		PY2 CN		N		2		
Social Skills Rating System					25					*	*	See app	Y												Y				
Strengths and Difficulties Questionnaire			N		7								Y				Y	Y		**			Y1	Y					
Timberlawn Child Functioning Scale													• Y							*									
Quality of Life Measures																													
Hopelessness Scale for Children								2		N	*	*							Y	**								Y	
Perceived Life Satisfaction Scale										*	****	**							Y-**D**	**	See app		Y1						
Quality of Life Profile: Adolescent version	Y											*			Y			Y		**								Y	
Students' Life Satisfaction Scale	Y; see app									Y	***	***	• Y						Y	**	*							N	

Measure	Cultural Sensitivity: Evidence of bias	Cultural Sensitivity: Cultural norms	Cultural Sensitivity: Used in Australia	Suitability: Cost	Suitability: Time	Suitability: Training	Suitability: Degree / experience	Suitability: Reading level	Suitability: Computer software available	Reliability: Test-retest within 2 weeks	Reliability: Test-retest coefficient	Reliability: Internal consistency	Reliability: Cross-informant agreement	Reliability: Interrater agreement	Content Validity: Theory-based items	Content Validity: Judgments	Content Validity: Rules	Content Validity: Respondent feedback	Construct Validity: Factor analysis	Construct Validity: Convergent validity	Construct Validity: Divergent validity	Construct Validity: Social desirability	Construct Validity: Group differences	Construct Validity: Sensitivity / Specificity	Norms: Current	Norms: Number of normative samples	Developmental sensitivity: Forms	Developmental sensitivity: Age effects	Developmental sensitivity: Norms
Safety Measures																													
Feelings of Safety items	Y																			*								Y	
Satisfaction Measures																													
Satisfaction Scales	N				10; see app							*; see app			Y				Y	*									
Youth Satisfaction Questionnaire					“few”							**																	
Self-Esteem Measures																													
Canadian Self-Esteem Inventory for Children										Y	****																		
Coopersmith Self-Esteem Inventory	Y				15					N	*	**							Y	*		*			Y	1	2	Y	Y
Culture-Free Self-Esteem Inventories II	Y			166	20		dd&	4	C* At*	C**** At****	C* At*									**	***				Y	1	2	Y	
Global Self-Worth Scale											**																		
Lipsitt Self-Concept Scale									*	*	***									*	*	*							
Multidimensional Self-Concept Scale	N			172	20						****														Y	1		N	
Perceived Self-Worth Questionnaire											**									*									
Personal Attribute Inventory for Children (PAIC and NPAIC)	Y; see app										***									**			see app			2; see app			Y

Measure	Cultural Sensitivity: Evidence of bias	Cultural Sensitivity: Cultural norms	Cultural Sensitivity: Used in Australia	Suitability: Cost	Suitability: Time	Suitability: Training	Suitability: Degree/experience	Suitability: Reading level	Suitability: Computer software available	Reliability: Test-retest within 2 weeks	Reliability: Test-retest coefficient	Reliability: Internal consistency	Reliability: Cross-informant agreement	Reliability: Interrater agreement	Content Validity: Theory-based items	Content Validity: Judgments	Content Validity: Rules	Content Validity: Respondent feedback	Construct Validity: Factor analysis	Construct Validity: Convergent validity	Construct Validity: Divergent validity	Construct Validity: Social desirability	Construct Validity: Group differences	Construct Validity: Sensitivity/Specificity	Norms: Current	Norms: Number of normative samples	Developmental sensitivity: Forms	Developmental sensitivity: Age effects	Developmental sensitivity: Norms
Self-Esteem Measures *cont'd*																													
Piers-Harris Self-Concept Scale	Y	Y		157	20			3		*	*	**					Y	Y	Y	**	***					2			Y
Rosenberg Self-Esteem Scale					5							**								***	*		N		N	1		N	
Self-Appraisal Inventory	Y									*	*	*							Y	**			Y						
Self-Description Questionnaire I	Y				20							*								**								Y	
Self-Description Questionnaire II					20							***							Y	*								N	
Self-Description Questionnaire III			Y		30					N	***	**													Y	1		Y	Y
Self-Esteem Index				245	30		dd&			*	***	***							see app	see app	see app	see app	see app	see app	Y	1		N	Y
Self-Perceived Role Competence	N									N	*	*								*									
Self-Perception Profile for Adolescents			Y									*							Y	*	***	*							
Self-Perception Profile for Children	Y	Y								N	**	*	Y						Y	**					Y	2		Y	
Tennessee Self-Concept Scale				293	20			2	Y	*	***	****								**; see app			Y		Y	2	2		
What I Think About Myself					10						***																Y		

Measure	Cultural Sensitivity: Evidence of bias	Cultural Sensitivity: Cultural norms	Cultural Sensitivity: Used in Australia	Suitability: Cost	Suitability: Time	Suitability: Training	Suitability: Degree/experience	Suitability: Reading level	Suitability: Computer software available	Reliability: Test-retest within 2 weeks	Reliability: Test-retest coefficient	Reliability: Internal consistency	Reliability: Cross-informant agreement	Reliability: Interrater agreement	Content Validity: Theory-based items	Content Validity: Judgments	Content Validity: Rules	Content Validity: Respondent feedback	Construct Validity: Factor analysis	Construct Validity: Convergent validity	Construct Validity: Divergent validity	Construct Validity: Social desirability	Construct Validity: Group differences	Construct Validity: Sensitivity/Specificity	Norms: Current	Norms: Number of normative samples	Developmental sensitivity: Forms	Developmental sensitivity: Age effects	Developmental sensitivity: Norms
Social Support Measures																													
Perceived Social Support Scale from Family (PSS-FA), Perceived Social Support Scale from Friends (PSS-FR); Perceived Social Support from Teachers (PSS-FT)												***							Y	**									
Personal History Inventory for Children										N	*									*									
Social Support Appraisals Scale: Revised	Y											***							Y	*									
Social Support Scale for Children and Adolescents												**								**									
Supportive Parenting Scale												**	• Y							*									
Stressful Event Measures																													
Adolescent Perceived Events Scale					See app					Y	***		• Y	***													3		
Daily Hassles Questionnaire												*****						Y		*	**								
Family Inventory of Life Events												*	• Y																
Hassles Scale for Children										Y	***	***																	
Life Events Checklist					15															*	**								
Sandler and Block's Modification of the Coddington Life Events Scale for Children	Y									N	*				Y			Y		*					Y				

Measure	Cultural Sensitivity: Evidence of bias	Cultural Sensitivity: Cultural norms	Cultural Sensitivity: Used in Australia	Suitability: Cost	Suitability: Time	Suitability: Training	Suitability: Degree / experience	Suitability: Reading level	Suitability: Computer software available	Reliability: Test-retest within 2 weeks	Reliability: Test-retest coefficient	Reliability: Internal consistency	Reliability: Cross-informant agreement	Reliability: Interrater agreement	Content Validity: Theory-based items	Content Validity: Judgments	Content Validity: Rules	Content Validity: Respondent feedback	Construct Validity: Factor analysis	Construct Validity: Convergent validity	Construct Validity: Divergent validity	Construct Validity: Social desirability	Construct Validity: Group differences	Construct Validity: Sensitivity / Specificity	Norms: Current	Norms: Number of normative samples	Developmental sensitivity: Forms	Developmental sensitivity: Age effects	Developmental sensitivity: Norms
Therapeutic Process Measures																													
Adolescent Working Alliance Inventory												***	• Y		Y		Y		Y-**D**	*									
Barriers-to-Treatment Participation Scale												*	• Y					Y	Y				Y4						
Family Involvement items	Y											***								**			Y1						
Loyola Child Psychotherapy Process Scales					See app									**					Y				Y1						
Motivation to Change Questionnaire																				**									
Penn Helping Alliance Questionnaire												****								*			Y2						
Psychotherapy Process Inventory modified for use with children						See app						***		K***	Y					*									
Smith-Acuna, Durlak, and Kaspars (1991) unnamed scales					15	See app						*	Y		Y		Y		N										
Therapeutic Alliance Scales for Children												C* CI**	• Y		Y		Y	Y		*									
Therapy Process Checklist												*			Y				Y	*; see app									
Treatment Acceptability Questionnaire										N	***; see app	***; see app			Y					***			Y1						
Treatment Evaluation Inventory	Y				4			5				***					Y		Y-**D**	***			Y2 or more						

Identifying the "Best" Measures

We do not claim that the detailed descriptions of the instruments included in this review are exhaustive. The results of the review are extensive and difficult to comprehend. The tables presented above do not clearly indicate whether the 188 instruments can be used to form the basis of a comprehensive measurement system for children and adolescents in Australia. To further refine the selection process, more exacting criteria were applied. Two raters examined each of the 188 measures within 18 of the 29 domains, according to the evaluation criteria listed below:

Evaluation Criteria Applied to the 188 Selected Measures

Asterisks are placed against those criteria that are of greatest importance.

1. **No evidence of cultural bias.** The measure (instrument) was examined for different psychometric properties with different cultural groups such as reliability or different means for each cultural group. If separate cultural norms exist, it may be possible to correct for cultural bias.

2. **Used previously in Australia.** Since the meaning of measures can vary with context, we examined whether measures had been tested in Australia.

*3. **Takes less than 10 minutes to complete.** Clinicians and clients have only a limited amount of time to complete measures. The 10-minute limit was set for single dimension instruments. For comprehensive instruments we set a maximum of 30 minutes. For some measures, time-to-complete is not indicated in any published information and the formula described earlier was used.

4. **Cost less than 25 cents per client.** This criterion keeps the costs down relative to installing a nationwide system. Very few proprietary measures meet this criterion, effectively limiting potential measures to those in the public domain.

*5. **No or less than 2 hours of required training.** This criterion is necessary because of the logistics and costs involved in extensive training of staff on a nation-wide basis. However, this criterion does limit potential measures to a fairly simple level.

*6. No degree or bachelor's degree required. This criterion facilitates use in clinical settings, as clinics may want non-professional staff to administer measures in order to save time and money. However, this criterion limits the complexity of potential measures.

7. Suitable reading level for specified age. Some instruments have high reading levels, negating their use as self-report measures for some of the target population. Furthermore, if the person reading the measure has to explain to the client what is meant by the question, this may change the characteristics of the measure.

8. Test-retest reliability within two weeks time and reliability greater than .70. We tried to focus on test-retest within 2 weeks because, as the time period increases, many more variables can intervene to change responses. The criteria for acceptable test-retest reliability ratings were based on Cross and McDonald (1995), Hammill, Brown, and Bryant (1992), and Rossi and Freeman (1993).

9. Internal reliability greater than .80. This is the degree to which individual questions on a scale measure the same thing. If it is low, two subjects may receive the same score but those scores could not be interpreted as meaning the same thing. The criteria for level of acceptable internal consistency ratings were based on Nunnally (1978).

10. Interrater reliability (where relevant) greater than .80. This is the degree of agreement or consistency among similar types of observers of the same phenomenon. It is indicated by such statistics as the Pearson correlation coefficient, the intraclass correlation coefficient, and Cohen's Kappa. We did not treat cross-informant agreement as equivalent to interrater reliability since we acknowledge that different informants do not necessarily observe the same phenomena and behaviour can vary greatly in different contexts. The criteria for acceptable interrater agreement ratings were based on Orwin (1994) and Nunnally (1978).

11. Content validity theory. We considered it very important that the measure's items be generated according to theoretical specifications of the content of the construct. Although not listed as necessary for the ideal measurement system, we examined other means of content validation: expert ratings of relevance, representativeness, or quality of items; use of rules or algorithms to generate or exclude items; and use of subject

matter experts (e.g., children, parents, teachers and clinicians) to generate or exclude items.

12. **Construct validity - factor analysis.** Factor analysis represents a variety of methods for discovering if test items collect into meaningful groups. We examined whether factor analytic studies had supported or disconfirmed the theorized structure of the measure.

*13. **Construct validity - Convergent validity greater than .50.** This is the degree to which a measure is related to a similar measure or criterion that represents theoretically the same or a similar construct. The criteria for level of acceptable convergent validity ratings were based on Cross and McDonald's ratings of levels of association (1995).

14. **Construct validity - Divergent validity less than .25.** This is the degree to which a measure is not related to another measure or criterion that it is not supposed to be related to. The criteria for acceptable convergent validity ratings were based on Cross and McDonald's ratings of levels of association (1995). When possible we also examined social desirability as a potential bias.

*15. **Construct validity - Known groups.** For a measure to be useful and valid, it must differentiate between groups that are theoretically different relative to the construct. A measure with known groups validity can be used diagnostically. Where available, we have recorded information on a measure's sensitivity and specificity.

16. **Current norms.** Current norms are important for clinical use as they provide expected scores that are representative of the overall population or sub-population. Norms provide the means for a clinician to compare a client's scores against a standard that indicates whether the client's scores are normal or atypical. We considered norms to be acceptable only if they were developed within the last 20 years.

17. **Developmental sensitivity.** A measure should take development into account. What may be appropriate and normal behaviour at one stage may not be appropriate and normal behaviour at another. When a measure does not take development into account, it has often been found that its psychometric properties vary with age. Some measures have alternate forms by age to deal with developmental differences.

18. **Developmental sensitivity - norms.** If a measure is not developmentally sensitive (i.e., the means vary by age), different norms by age are a corrective.

No measure met all 18 criteria. The criteria were too strict. Within each domain, the raters then evaluated those measures satisfying the most important criteria (the starred items 3,5, 6,13,15). Ten measures met these five criteria. These measures were included in the “best measures” list. Since even the reduced number of criteria eliminated so many measures, we applied additional judgment. Within each domain, raters subjectively weighted certain criteria to select the best measures within that domain. First priority was given to measures demonstrating convergent validity. If this was not present then the next priority was presence of group difference evidence. If a domain lacked any measures with acceptable validity data, we searched for measures with acceptable reliability. Across all domains an attempt was made to pick measures that met time feasibility requirements. Using these criteria, raters came to 97% agreement (Kappa = .89) on the 15 “best measures” in each domain. We then added those 6 measures to the “best measures” list that had been selected only by one rater.

In each domain, those instruments that were selected for the “best measures’ list received a further literature search. We specifically searched for measures in which starred-criteria data were missing. We searched the measures with no limitations (no delimiters) and recovered all articles that were available at the Vanderbilt library). Interestingly, the additional articles added little new information. We believe that the reviews of measures, from which a great deal of our information came, effectively covered their psychometric properties.

The 31 measures listed in Table IV.2 are also described in some detail in the text of the report. The reader can examine Table IV.2 for an overview and to compare the qualities of instruments within and among domains. The text material is easier to access for descriptions of specific instruments.

Special mention needs to be made about the absence of the Child Behaviour Checklist (CBCL) from the best 31 instruments list. The CBCL is a 118-item parent and youth reported measure. One of the most widely used general symptom measures, it has been extensively researched. The CBCL measures problem behavior in children over the past 6 months. It has externalizing and internalizing scales and it also includes a number of subscales: (e.g., withdrawn, somatic complaints, anxious/depressed, social problems, thought problems, attention problems, delinquent behavior, aggressive behavior, and sex problems). The CBCL has highly acceptable convergent validity and effectively discriminates between at least five different clinical groups.

Nonetheless, the CBCL did not make the "best measures" list. The CBCL was compared to measures in the domain of general symptoms, a domain containing several acceptable measures (i.e., the BDRS, the PSC, and the VSTCP). The primary reason the CBCL did not make the best measures list was its length, given the need to administer other instruments in a composite measure. The CBCL takes 15-17 minutes to complete. The other three measures take ten minutes or less. Furthermore, the CBCL problems scale is supposed to be administered along with the CBCL social competence making the measure even lengthier. The YSR (the youth version of CBCL) has unacceptable two-week test-retest reliability, and the CBCL has unacceptable internal consistency. In addition, the CBCL is a proprietary instrument, which would make it very expensive for Australia to adopt on a widespread basis. Finally, recent research in Australia has found inconsistent support for the norms developed in the United States for the CBCL (Achenbach, T. M., Hensley, V. R., Phares, V., & Grayson, D. ,1990; Bond, L., Nolan, T., Adler, R., & Robertson, C., 1994; Hensley, V. R. ,1988).

The "best measures" list represents the results of a comprehensive literature review and the application of specified quality criteria. The next step in the evaluation is to determine if some combination of instruments on this list would meet the criteria of an ideal measurement system. If the combination of instruments falls short, we will consider what would be needed to produce an ideal, feasible measurement system.

Results

A Critical Review of Those Instruments that Best Satisfy Selection Criteria

Table IV.2 — Measures that Best Satisfy Selection Criteria

Domain	Measure	Page
Coping Skills	Coping Strategies Inventory	106
Family Functioning	Family APGAR	107
	Family Environment Scale	107
	Structural Family Interaction Scale-Revised	109
Family Resources	Family Resource Scale	110
Functional Competence	Behavioural and Emotional Rating Scale	111
	CBCL Social Competence Scale	112
Functioning Impairment	Children's Global Assessment Scale	113
	Columbia Impairment Scale	113
General Symptoms	Behaviour Dimensions Rating Scale	114
	Pediatric Symptom Checklist	115
	Vermont System for Tracking Client Progress	115
Goals	Family Goal Recording	116
Maltreatment	Child Abuse Potential Inventory	116
	Child Report of Treatment Issue Resolution	117
Multidimensional	Adolescent Treatment Outcomes Module	118
	Behaviour Assessment System for Children	118
	Ohio Youth Problems, Functioning, and Satisfaction Scales	119
	Social Adjustment Inventory for Children and Adolescents	120
Quality of Life	Perceived Life Satisfaction Scale	121
	Students' Life Satisfaction Scale	122
Satisfaction	Satisfaction Scales	123
Self-Esteem	Personal Attribute Inventory for Children	124
	Rosenberg Self-Esteem Scale	125
	Self-Appraisal Inventory	126
Social Support	Perceived Social Support Scale from Family/ Friends; adapted Perceived Social Support Scale from School Personnel	126
Stressful Events	Daily Hassles Questionnaire	127
	Hassles Scale for Children	128

Table IV.2 — Measures that Best Satisfy Selection Criteria cont'd

Domain	Measure	Page
Therapeutic Process	Family Involvement Items	128
	Treatment Acceptability Questionnaire	128
	Treatment Evaluation Inventory	129

Coping Skills

Coping Strategies Inventory (CSI) (Tobin, Holroyd, & Reynolds, 1984) : The CSI is a 72-item, adolescent self-report (five-point, Likert-type format) that has been tested with high school and college student populations. Of the 72 items, 23 items are derived from the Ways of Coping Scale (Folkman & Lazarus, 1985). The CSI measures 8 types of situational coping skills: *problem solving*, *cognitive restructuring*, *emotional expression*, *problem avoidance*, *social support*, *wishful thinking*, *social withdrawal*, and *self-blame*. The reading level is unknown, but the measure was originally designed for use with adolescents and young adults. Based on our formula, the CSI requires 11 minutes to complete. The measure is in the public domain. Except for basic test-administration knowledge, it requires no special training or experience to administer. Two-week test-retest reliability across subscales ranged from .67 to .83, which was considered barely acceptable. Internal consistency (Cronbach's Alpha) was .71 to .94 across subscales (also barely acceptable). Content validity was supported through factor analysis and inclusion of items with the highest factor loadings. Tobin, Holroyd, Reynolds, and Wigal (1989) and Cook and Heppner (1997) each replicated the original factor structure. Studies of coping skills in individuals with different life problems showed that the CSI differentiated the following groups of individuals: depressed and non-depressed; neurotic and non-neurotic; and individuals with and without headaches. Convergent validity ranged from .33 to .77 for those 8 of 10 Kidcope subscales hypothesized to relate most closely with the CSI subscales. We consider this convergence unacceptable, since the correlation of .33 is low. Normative data and data about developmental and cultural sensitivity have not been reported. To our knowledge, computer software has not been produced for this measure.

Family Functioning

Family APGAR (Smilkstein, 1978) : The Family APGAR is a 5-item, parent or child self-report that uses a 3-point, Likert-type scale. This instrument is a brief screening tool that measures *family adaptation, partnership, growth, affection*, and *resolve*. The reading level is unknown. The measure was originally designed for adults, but has been used with children as young as 10 years of age. The test should take only an estimated five minutes to complete. The Family APGAR is in the public domain. Except for basic test administration knowledge, it should require no special training or experience to administer. Two-week test-retest reliability is mostly acceptable (.83). Internal consistency (Cronbach's Alpha) ranged from .80 to .86, which we rate as barely to somewhat acceptable. The theory-based nature of the items supports the measure's content validity, although it is difficult adequately to cover the important dimensions of family functioning with only 5 items. Convergent validity was highly acceptable: the Family APGAR correlated .80 with the Family Functioning Index. The measure significantly differentiated clinical and normal groups. Sensitivity and specificity in identifying adult patients with psychological distress were 68% and 62% respectively, with 16% false positive and 19% false negative rates. Although it is difficult to evaluate identification rates, these percentages appear to be unacceptable. Clinical cutoffs are suggested but there are no normative data. Developmental sensitivity is unknown. In one study, significant differences in scores were found across levels of parental education. To our knowledge, computer software has not been produced for this measure.

Family Environment Scale (FES) (Moos & Moos, 1981): The FES is a 90-item self-report that can be filled out by any family member. The FES measures family social environment and can include measuring level of agreement among family members. The measure taps 10 sub-domains within three broad dimensions: *relationships, personal growth*, and *system maintenance*. The FES can be administered in three forms: *Real, Ideal*, and *Expectations*. The Real Form assesses an individual's actual perceptions; the Ideal Form asks the individual to describe the ideal family environment; and the Expectations Form assesses an individual's expectations about family settings. The FES reportedly only takes 10 minutes to complete, although extra time would be required to assess multiple family members or administer multiple forms. Reading levels are not known, but presumably the availability of child and adult forms makes it appropriate for most or all reading levels. The FES costs

$44 for the scoring manual, $101 for 100 forms, $50 for 100 answer sheets, and $12 for the scoring key. Except for basic test administration knowledge, the measure requires no special training or experience to administer. Two-month test-retest reliability ranged from .68 to .86 across sub-scales, which indicates barely acceptable retest reliability. Internal consistency (Cronbach's Alpha) ranged from .10 to .89 across subscales and studies, which indicates the unacceptable internal reliability of certain subscales in certain settings. Content validity was established through theory generation (for the three dimensional constructs), and item formulation was based on observations and interviews with families, and rules for wording and inclusion of items. However, the measure's content validity comes into question considering that in two samples of graduate student raters, two-thirds of one sample placed only 24 of 45 items in correct subscales and two-thirds of another sample placed only 39 of 45 items correctly. According to Cross and McDonald's (1995) review, a confirmatory factor analysis study did not support the already established subscales. Furthermore, the only two factors to emerge explained up to 34% of the variance.

Studies have found that the FES varies in the predicted direction according to family composition and presence of specific problems in family. Convergence with other theoretically-related measures has covered a wide range: (a) correlations ranging from .17 to .73 with the Structural Family Interaction Scale; (b) a .89 correlation of the FES cohesion scale with the Structural Family Interaction Scale-Revised (SFIS-R) (enmeshment-disengagement subscale); (c) a correlation of -. 59 of the FES control scale with the SFIS-R (flexibility-rigidity subscale); (d) .78 and .68 correlations of the FES cohesion scale with the Family Assessment Device (FAD) (problem-solving and "AFFIN" subscales); (e) a .86 correlation of the FES cohesion subscale with the Family Adaptation and Cohesion Evaluation Scale-III (FACES-III) (cohesion subscale). Skinner (1987) reported that, in one study, the FES did not correlate with other measures of family cohesion. Using the lowest correlation (-.59) with the measure that had the most highly correlated subscale (.89 with the SFIS-R enmeshment-disengagement subscale), we rated the measure's convergent validity as barely acceptable. Divergent validity was rated unacceptable due to moderately high correlations of FES subscales with theoretically non-related dimensions from other measures (SFIS-R, FACES III) (ranging from .23 to .65). Social desirability was also rated unacceptable as certain subscales were moderately correlated with desirability scales (range from

.02 to .44). Fairly extensive evidence of the measure's ability to differentiate various groups (such as treated versus untreated families) provides some evidence of its validity. However, the FES discriminated between high-service-use families and low-service-use families on only 5 of 10 subscales. It is arguable whether this finding supports the measure's validity since service use is not necessarily highly related to family functioning.

Two large samples have provided normative data for normal and distressed families, size of family, single parent families, adults and adolescents, and by ethnic background (African American, Mexican American, Caucasian). According to Cross and McDonald (1995), the FES is reliable with samples of African Americans and Hispanics but has very low reliability with Puerto Ricans or Vietnamese refugee youth. The measure may not translate well due to dissimilar value orientation, colloquialisms, or negatively worded questions. Developmental sensitivity is unknown, although a separate version, published in 1984, is available for children. Hammill, Brown, and Bryant (1992) reported a test-retest less than .80. Child norms were collected sometime between 1976 and 1985. To our knowledge, computer software has not been designed for the FES.

Structural Family Interaction Scale-Revised (SFIS-R) (Perosa, Hansen & Perosa, 1981): The SFIS-R parent self-report is a 68-item, 4-point scale that identifies family interactions as described in Minuchin's (1974) structural model of family functioning. The revised version is based on a factor analysis of the original scale (Perosa, Hansen, & Perosa, 1981). The 7 SFIS-R subscales include *enmeshment-disengagement* (EN/D), *flexibility-rigidity* (FLX/RG); *family conflict avoidance-expression* (FCA/E); *mother-child cohesion-estrangement* (MCC/E), *father child cohesion-estrangement* (FCC/E), *spouse conflict resolved-unresolved* (SPCONR/U), and *cross-generational triads-parent coalition* (CGT/PC). Based on our formula, the SFIS-R requires 11 minutes to complete. The availability of the measure, and the training and experience needed for administration are unknown. Four-week test-retest reliability with college students ranged from .81 to .92 across subscales, which was considered mostly acceptable. Cronbach's alpha coefficients have ranged from .71 to .94 across subscales, which we rate barely acceptable. Convergence with the FES, FACES III, and FAD has yielded a wide range (see above description of the FES). SFIS-R correlations with the latter two measures include (a) .85 (EN/D scale) with FACES III (cohesion subscale); (b) .40 correlation (EN/D) with FACES

III (adaptability subscale); and (c) .49 (FCC/E) with FAD (AFF IN). As with the FES, we rate the measure's convergent validity as barely acceptable. Divergent validity is rated unacceptable due to moderately high correlations of SFIS-R subscales with theoretically non-related dimensions from other measures (e.g., FES, FACES III, and FAD) ranging .15 to .70. The measure has distinguished normal families, clinical family controls, and families in which incest had occurred, and has differentiated family variables associated with clinical populations from those related to controls. To our knowledge, normative, developmental sensitivity, and cultural sensitivity data have not been reported.

Family Resources

Family Resource Scale (FRS) (Dunst & Leet, 1987) : The FRS is a 30-item, parent self-report. The measure uses a 5-point, Likert-type scale with a "does not apply" option. Items are ordered roughly from the most basic needs (most important) to the least basic. The reading level is unknown, but the measure is designed for parents of young children. Based on our formula, the FRS requires 7 minutes to complete. The measure is in the public domain, and except for basic test administration knowledge, should require no special training or experience to administer. The FRS measures parental perceptions of the adequacy of different family and household resources (e.g., time, energy, money, and adequacy of food, clothing, and shelter). Test-retest (2 to 3 months) reliability was .52, which is unacceptable. However, adequacy of various resources could fluctuate over a period of 2 or 3 months, so a shorter test-retest period would be preferable. Internal consistency (Cronbach's alpha) and split-half reliabilities were mostly or highly acceptable (.92 and .95, respectively). Evidence of content validity is that items were theory-based and rank-ordered by professionals, with a .81 coefficient of agreement. Correlations with measures of socioeconomic status, parent well-being, parent commitment to intervention, informal social support, mother's commitment to child care, and frequency of social contact ranged from .39 to .62, most of which levels of convergence were considered unacceptable. Normative, developmental, and cultural sensitivity data have not been reported. To our knowledge, computer software has not been produced for this measure.

Functional Competence

Behavioral and Emotional Rating Scale (BERS) (Epstein & Sharma, 1998): The BERS is a 52-item self-report completed by adults who know the child or adolescent, including parents, family members, teachers, or professionals. The BERS uses a 4-point scale ranging from "not at all like the child" to "very much like the child." The BERS measures behavioral and emotional strengths manifest within the past 3 months with an overall strength quotient and five subscales: *interpersonal strengths*, *family involvement*, *intrapersonal strengths*, *school functioning*, and *affective strength*. Eight open-ended items are also available, if required. The reading level of the measure is unknown. The measure requires 10 minutes to administer, and basic training in test administration and scoring. The cost is $74 for the introductory kit (and manual) and $34 for every 50 forms (68 cents each). Test-retest reliability (within 2 weeks) among 59 students aged 14 to 19 years was mostly acceptable to highly acceptable (.85 to .99 across subscales and .99 on the overall strength quotient). Internal consistency averages for subscales in two normative samples ranged from .84 to .96 (overall strength quotient was .97 and .98), which we rated somewhat to highly acceptable. Interrater reliability between 9 pairs of special education teachers was .83 to .96 across five subscales and .98 on the overall strength quotient, mostly acceptable to highly acceptable levels of agreement. Interrater scoring reliability on 30 measures was an excellent .99 for one pair of raters. Content validity was achieved with the following steps: (a) 250 professionals were asked to generate behaviours (1,200 behaviours were identified and 190 were found non-redundant, in 15 categories); (b) 400 professionals rated the relevance of 190 items to child strengths (63 lower-scoring items were excluded); (c) 37 of 127 items did not discriminate between children with and without serious emotional disorder, and were dropped while 10 items were excluded due to their infrequency; (d) factor analysis resulted in a 68-item, 5-factor scale; (e) using an item discrimination validity cut-off of .3, 16 items were deleted; (f) item discrimination validity of the 52-items scale was computed with the two normative samples, ranging from .38 to .80 across subscales and age. Convergent validity was tested by correlating the BERS with the Walker-McConnell scales (.29 to .85), Self-Perception Profile for Children subscales (.28 to .72), and the Teacher Report Form (.28 to .75). These validity coefficients were deemed unacceptable due to the low correlations (below .50) for each comparison measure. The BERS significantly distinguished between children with and children without emotional and behavioural disorders on each scale (children with disorder scored lower on strengths). Norms were established using a nationally representative, standardized sample of 2,176 students aged 5-18 years and a separate group of 861 children with emotional or behaviour disorders. Separate norms are available by age, gender, and ethnicity. Although developmental and cultural sensitivity data have not been reported, the published norms allow the user of the instrument to compare results to the relevant norms. Computer software is not available for the BERS.

Child Behavior Checklist and Youth Self-Report Social Competence Scales (CBCL-SCS) (Achenbach, 1991a) (YSR-SCS) (Achenbach, 1991b) : The social competence subscales of the CBCL and YSR are parent and youth reports of the youth's social competence over the past 6 months across three subscales: *activities*, *social*, and *school*. The CBCL-SCS has 20 items; the YSR-SCS has 14 items. The response format varies with the different domains. The scale is suitable for those with at least a 5th grade reading level. The SCS requires 5-10 minutes to complete. The measure is not in the public domain, and costs $143 for the introductory kit and 100 forms. No training is required to use this measure. For the CBCL-SCS, one-week test-retest reliability (ICC) was found to be a mostly acceptable .87 to .99. For the YSR-SCS, test-retest over an unspecified time period was found to be at a somewhat acceptable mean of .80. Internal consistency for both informants is an unacceptable .42 to .64 across subscales and studies. Interrater reliability, when used as an interview instrument, was a highly acceptable .93. Inter-parent reliability was found to be a mean of .79. When the measure was factor-analyzed the activities and school subscales were found not to load on a social competence factor. The SCS has barely acceptable convergent validity as determined by a correlation of .52 with the SAICA. The SCS has been found to discriminate significantly between clinic-referred and control youths. It has current norms by gender and for ages 4-11 and 12-18 for the CBCL and 11-14 and 15-18 for the YSR. The YSR-SCS is developmentally biased, with stronger test-retest reliability for older adolescents. Statistically it has not been found to be culturally biased but it has been criticized anecdotally as insensitive to the functioning of youths in impoverished or dysfunctional families. Computer software is available for this measure. It should be noted that the SCS is a subscale of the CBCL and must be used with the full instrument unless special permission is obtained from the author.

Functional Impairment

Children's Global Assessment Scale (CGAS) (Shaffer et al., 1983): The CGAS is a single item scale rated from 1 to 100 by a clinician or lay interviewer after gathering information from the child and/or parent. The CGAS measures global functioning with one overall score, using behaviour-oriented descriptions for every 10 points. A score of less than 70 is said to indicate a "case". Reading level is not an issue as this is a clinician rating scale. Time-to-complete varies with the amount of information the clinician has about the child. The measure is in the public domain, and Bird and associates (1987) claim minimal training is needed. The measure does not come with formal training materials. Test-retest reliability has often been studied but the time period has rarely been specified. This makes it difficult to evaluate the meaning of the results. Test-retest reliability was found to be a mostly acceptable .85 on case vignettes over a period of six months. A much shorter time period with actual clients would be preferable. Internal consistency does not exist, as the CGAS is a single-item measure. There have been several multiple informant studies with the CGAS. For example, agreement between attending psychiatrists and milieu staff has been found to range from (Pearson r) .62 to.76. Several studies with traditional mental health clinicians have produced barely acceptable interrater intraclass correlations (from .53 to .93). The CGAS has highly acceptable convergent validity as it has been found to correlate from -.82 to -.92 with Axis V ratings, -.76 to -.90 with psychiatrist severity ratings, and -.80 with the CGAS total problem score. The CGAS has been found to discriminate significantly between at least 4 different groups. Sensitivity and specificity have been examined many times and are usually greater than .80. The CGAS has not been found to be developmentally biased. Cultural sensitivity and normative data have not been reported. Computer software is not applicable to this one-item measure.

Columbia Impairment Scale (CIS) (Bird et al., 1993): The 13-item CIS assesses functional impairment on a five-point scale across the domains of *interpersonal relations*, *psychopathology*, *job or school functioning*, and *use of leisure time*. A lay person who interviews the parent or child completes the measure. The CIS requires 5 minutes for completion. No information was obtained from the literature concerning training requirements. The instrument is in the public domain. Test-retest reliability has been reported but the time period has not been specified which makes it difficult to evaluate the meaning of the results. Test-retest reliability was a mostly acceptable .89 (ICC) for parents and a somewhat acceptable .63 (ICC) for children. There were no data on internal consistency or multiple informant agreement. The CIS has barely acceptable convergent validity for parent report as it has been found to correlate .63 with the clinician CGAS. The CIS has unacceptable convergent validity for the child report as it has been found to correlate .43 with the clinician CGAS. The CIS has been found to discriminate significantly between at least two different groups. Normative, developmental and cultural sensitivity data have not been reported. To our knowledge, computer software has not been produced for this measure.

General Symptoms

Behavior Dimensions Rating Scale (BDRS) (Bullock & Wilson, 1989): The BDRS is a 43-item counselor-teacher-or parent-reported measure. The informant must have known the participant for at least 2 weeks. The BDRS screens for behavior patterns with a 7-point scale across the domains of *aggression*, *inattention*, *social withdrawal*, and *anxiety*. Information on reading level has not been reported in the literature, an important issue for a parent-report measure. Time to complete varies from five to 10 minutes. The measure is not in the public domain and costs $239 for the introductory kit and 100 forms. No training is required. Test-retest reliability has been reported but the time period has not been specified which makes it difficult to evaluate the meaning of the results. Test-retest reliability was a mostly acceptable .82 to .89. Internal consistency was a mostly acceptable .90 to .95 across subscales. However, interrater agreement was an unacceptable .60 to .68 between teachers and teacher assistants. For content validity, experts confirmed the subscale content on a 30-item version, confirmatory factor analysis establish adequate fit with the hypothesized structure. The BDRS has not been adequately validated and the manual is said to contain unsubstantiated claims (Vaidya, 1996). Multi-trait multi-method validation suggests good convergent validity only for the aggression and inattention scales. The BDRS has been found to discriminate significantly between behaviorally disordered and non-disordered children with 74 to 76% correct placements by subscale. The BDRS is developmentally biased and has separate norms by grade level. The BDRS has been standardized on 1,942 Canadian subjects, and has separate norms for 641 youth in correctional facilities. There are also separate norms by gender. Hammill, Brown, and Bryant (1992) rate the BDRS as having an unacceptably low percentage of females in the normative sample. Cultural sensitivity data have not been reported. We do not know whether computer software is available for this measure.

Pediatric Symptom Checklist (PSC) (Murphy & Jellinek, 1985): The PSC is a 35-item parent reported measure that can also be orally administered. It is a shortened and revised version of the Washington Symptom Checklist. The PSC is a 3-point screening measure of children's behavioral/emotional symptomatology. Information on reading level has not been reported. However, Jellinek and associates (1995, p. 742) report that the PSC is "well-liked by parents, praised by pediatricians, and accepted by clinic staff." The PSC is reported to have a 90% completion rate in clinics. The PSC requires 5 minutes to complete. The measure is in the public domain. No information is available on whether training is required to use this measure. Test-retest at one week was found to be a mostly acceptable .85 to .86 in two samples. Agreement on "caseness" at one week was .69 (kappa). Internal consistency was a somewhat acceptable .86 to .94 across studies, and interrater agreement a barely acceptable .52 to .82 (kappa). The PSC has barely acceptable convergent validity, correlating of .64 with the CBCL and .63 with the CGAS. The PSC has been found to discriminate significantly between 7 different groups. Sensitivity (.42 to .95), specificity (.68 to 1.00), and agreement (kappa = .52 to .82) on "caseness" have been examined many times and have varied greatly across studies. The PSC has no developmental sensitivity data but it has a different cutoff score for "caseness" in younger children. The PSC has current norms. The PSC is culturally biased as the case rate has been found associated with SES (r = -.12) and sensitivity and specificity are significantly different depending on SES level. We do not know whether computer software is available for this measure.

Vermont System for Tracking Client Progress (VSTCP) (Burchard & Bruns, 1993) : The VSTCP is a 22-item service provider-or primary caregiver-reported measure. The VSTCP measures *behavioral, emotional, life event,* and *educational* outcomes considered indicators of the risk for out-of-home placement. It consists of Daily, Weekly, Monthly, and Quarterly Adjustment Indicator Checklists, measuring 16 negative and 6 positive behaviors. It also includes a *restrictiveness of living environment* scale. Information on reading level has not been reported in the literature. Based on our formula, the VSTCP requires 6 minutes to complete. No information was obtained from the literature or catalogues as to the cost of this measure or any training requirements. Test-retest and interrater agreement data has not been reported. Internal consistency (Cronbach's alpha) was a barely acceptable .77 to .85 for negative behaviors. Factor analysis provides some evidence of validity; however, the factors have not been found to be clear across analyses. The VSTCP has barely acceptable convergent validity: the problem scale correlates greater than .60 with the CBCL, .50 with restrictiveness, and .50 to .59 with one-year case-management costs. The positive behaviours correlate -.54 with restrictiveness and -.70 with cost. The VSTCP has been found to discriminate significantly between a "wraparound" care group and a comparison group. Cultural and developmental sensitivity and normative data have not been reported. The measure has been studied only in Vermont, predominantly with white males. It may also be subject to case-manager biases in rating improvement. Computer software is available for this measure.

Goals

Family Goal Recording (FGR) (Fleuridas et al., 1990): The FGR, an adaptation of Goal Assessment Scaling, is a procedure used by a therapist to generate any number of goals. In order to operationalize and measure change related to target complaints. Reading level is not an issue as this is a clinician-rating scale. Time-to-complete is one session. No information was obtained from the literature or catalogues as to cost of this measure or any training requirements for use. No test-retest or internal consistency data has been reported. Interrater agreement has been found to be a highly acceptable kappa of .78 between raters and therapy teams. The FGR has barely acceptable convergent validity as it has been found to correlate -.62 with the mother's dissatisfaction with her children. However, the FGR is correlated only .22 with the father's dissatisfaction with the children. It is correlated -.68 with father's global distress with marriage, -.26 for mother's. Cultural and developmental sensitivity and normative data have not been reported. Computer software is not available for this measure.

Maltreatment

Child Abuse Potential Inventory (CAPI) (Milner, Gold, & Wimberly, 1986): The CAPI is a 160-item parent/caregiver self-reported measure. It is a screening device, using an agree-disagree format, aimed to differentiate physical abusers from non-abusers, with a 77-item physical abuse subscale and three validity scales. The CAPI is suitable for those with at least a third grade reading level, and requires 12-20 minutes to complete. No information was obtained from the literature or catalogues as to the cost of this measure or any training requirements for use. One-day test-retest reliability was a highly acceptable .91 (.90 at one week). Internal consistency was a barely acceptable .74 to .98 across subscales and studies; however, it was as low as .24 for one subscale of a Greek version of the measure. In support of the measure's content validity, the authors discuss its development from a comprehensive literature review of child abuse and neglect. In addition, many items were eliminated that did not significantly discriminate abusing from non-abusing parents. A later study found that 77 of the items significantly discriminated abusing parents. In a Greek version of the measure which was factor analyzed, the items did not all constitute the same factors as in the original factor analysis. A second factor analysis also found a different number of factors from those of the original factor analysis. The CAPI has highly acceptable convergent validity as seen by a correlation of .79 with parent global severity. It also correlates .55 with negative parenting style. The CAPI has been found to discriminate significantly between four different groups, with 77 to 93% correct classifications across studies. The CAPI has current norms on 836 adults. The CAPI has been found to be culturally and developmentally biased with significantly different scores depending on nationality, SES, marital status, and age of parents. We do not know whether computer software is available for this measure.

Child Report of Treatment Issue Resolution (CRTIR) (Nelson-Gardell, 1997): The CRTIR is a 38-item adolescent (10 to 17 years) self-reported measure that assesses the effectiveness of therapy intervention for sexual abuse victims. Information on reading level has not been reported, an important issue for a youth-report measure. Based on our formula, the CRTIR requires 8 minutes to complete. No information was obtained from the literature or catalogues as to the cost of this measure or any training requirements for use. No test-retest data have been reported. Internal consistency was mostly acceptable (alpha = .93). In support of the measure's content validity, experts rated individual item relevance. The CRTIR was factor analyzed into four factors explaining 47% of the measure's variance. The CRTIR has barely acceptable convergent validity as seen by a correlation of -.74 with the Trauma Symptom Checklist for Children. It is unacceptably biased toward socially desirable responses as it has been found to be correlated .45 with the Children's Social Desirability Questionnaire. The CRTIR has no normative data or data on cultural or developmental sensitivity. We do not believe computer software is available for this measure.

Multidimensional Scales

Adolescent Treatment Outcomes Module (ATOM) (Robbins et al., 1997) : The ATOM is a 68-item youth-and parent-self-reported measure. It assesses *symptom severity, functioning at home, school functioning, troubles with rules and laws*, *leisure functioning*, *friendships*, *dysfunctional peers*, *family burden*, and *satisfaction*. It is suitable for those with at least a 4th grade reading level. The ATOM requires 25-28 minutes to complete for youth and parents, 7 minutes for a clinician. It is copyrighted but its authors report that it can be used at no cost. Test-retest reliability has been reported, but since the time period has not been specified it is difficult to evaluate the meaning of the results. Test-retest reliability was found to be a barely acceptable .64 to .86 across subscales. Internal consistency was associated with an unacceptable alpha of .60 to .94 across scales and informants. In support of the measure's content validity, items were derived from review of the literature and a panel of experts, who suggested what is relevant for understanding how treatment affects outcomes. The ATOM has minimally acceptable convergent validity for parent report: focal problem severity is correlated (Spearman) .51 with the CBCL/YSR internalizing score; and ATOM symptom severity is correlated .69 with the CBCL/YSR internalizing score. Symptom severity is correlated -.74 with CHQ behaviour problems. The convergent validity of the ATOM youth report is unacceptable: there is only a .13 correlation of focal problem severity and only .43 correlation of symptom severity with the CBCL/YSR internalizing score. Symptom severity for youth report correlated only -.23 with CHQ behavior problems. Parent report of burden was correlated -.53 to -.55 with appropriate scales of the CHQ. The ATOM discriminated significantly among four different groups with .47 to .65 correct classification across diagnoses (using kappa), but did not differentiate ADHD from non-ADHD (kappa = .37). There are no normative data or data on cultural or developmental sensitivity. We do not believe computer software is available for this measure.

Behavior Assessment System for Children (BASC) (Reynolds & Kamphaus, 1992): The BASC is a 148-item measurement system, on a 4-point scale, consisting of a *Teacher Rating Scale* (TRS), a *Parent Rating Scale* (PRS), a *Self-Report of Personality* (SRP), a *Structured Developmental History Inventory*, and a *Student Observation System* (classroom). The BASC provides composite scores for *behavioral symptoms*, both externalizing and internalizing, *school problems*, and *adaptive skills*; the self-report composite scores include *emotional symptoms, school maladjustment, clinical maladjustment*, and *personal adjustment*. The BASC also has *validity* of *report* scales. Information on reading level has not been reported in the literature. The BASC requires 10 to 20 minutes for parents and teachers, 30 minutes for youths; and 15 minutes of observation to complete. The measure is not in the public domain, and, costs $125 for the introductory kit and 100 forms. It is unknown whether training is required to administer this measure. Two-to-8- week test-retest reliability was found to be a mostly acceptable .82 to .91 across subscales for teachers, and a barely acceptable .70 to .88 for parents. One-month test-retest reliability for the youth report was a somewhat acceptable .76 to .86. A shorter time period would be preferable to measure test-retest reliability. Internal consistency was somewhat acceptable (greater than .80 for all scales on teacher report). However, it was unacceptable for parent report, (.56 to .94), and for youth report, (.54 to .97). Across scales and ages, teacher-interrater reliability was an unacceptable .29 to .93 and for parents it was .46 to .67. In support of the measure's content validity, items were derived from review of the literature, from other measures, and from clinical experience. Items were developed in phases and scales were determined by factor analysis. The BASC has barely acceptable convergent validity (the parent report of externalizing behavior is correlated .67 to .78 with the Connors Parent Rating Scale); and the internalizing scale was correlated .45 to .51 with the Connors Parent Rating Scale. The BASC did not adequately discriminate between referred and non-referred samples. The BASC was community-normed between 1988 and 1991 at 116 testing sites with 2,401 teachers, 3,483 parents, and 9,861 children. Clinical norms were derived from 36 sites involving: 693 teachers, 401 parents, and 411 children. Gender and ethnicity-biased items were dropped during development. Separate norms are available by gender. There are separate versions by age of the youth. There may be developmental bias as there is a higher correlation of parent and teacher reports with older youths. No cultural bias has been detected. Computer software is available for this measure.

Ohio Youth Problems, Functioning, and Satisfaction Scales (OYPFSS; Ogles, Davis & Lunnen, 1998): The Ohio Scales measure *problem behaviours*, *current functioning* (within the past 3 months), and *satisfaction with services*. The measure is designed to be sensitive to changes in client strengths and weaknesses during intervention (Ogles, personal communication, July 15, 1997). The scales are self-reports completed by parents, youth, or case managers. The OYPFSS consists of 44 problem items, 20 *functioning* items, one *satisfaction* item, 3 *inclusion* items, 4 *hopefulness and well-being* items, and 2 *family functioning* items. Some domains vary by informant, although each domain uses a Likert-type format. The reading level is unknown. The first author stated that the scales require minimal training to administer and 8-10 minutes to complete. The measure can be obtained from the authors at no cost. Test-retest reliability has not been reported. Internal consistency was rated somewhat acceptable for parents (.76 to .97) and youth (.74 to .96), but unacceptable for case manager respondents (.44 to .94). The measure was developed through a review of research studies, a conceptual development process to determine areas relevant to clinical outcomes assessment, and a stakeholder survey to identify areas deemed most important. Items were based on DSM-IV, a community mental health center list of common presenting problems, a review of instruments, a survey, and consultation with service providers. The parent scales correlate .89 (*Problems*) and .77 (*Functioning*) with the Child Behaviour Checklist. The youth scales correlated .82 (*Problems*) and .46 (*Functioning*) with the Youth Self Report. The case manager scales correlated -.47 (Problems) and .38 (Functioning) with the Progress Evaluation Scales (PES), while change in problems correlated -.54 with changes in the PES and change in functioning correlated - .56 with changes in the PES. These convergent validity coefficients are highly acceptable (parent), unacceptable (youth) and barely acceptable (case manager). Evidence of group difference validity was found as four of five groups (paired comparisons) had significantly different means on symptoms, functioning or both. Normative data include clinical and community samples recently studied in southeast Ohio. Clinical cutoffs are based on *t*-scores. Norms are not published separately by age. Computer software is not available for the Ohio Scales.

Social Adjustment Inventory for Children and Adolescents (SAICA) (Gammon et al., 1982): The SAICA is a 77-item semi-structured interview, using a 4-point scale, of the youth and a parent. It provides systematic assessment of children's adaptive functioning in school and spare-time activities, with peers, siblings, and parents; it has 35 *competence* items and 42 *problem behavior* items. Information on reading level has not been reported in the literature and is not likely to be relevant to a semi-structured interview. The interview has been found acceptable to both parents and children. The SAICA requires 30 minutes to complete. No information was obtained from the literature or catalogues as to the cost of this measure. The SAICA is to be administered by an interviewer knowledgeable in child development and with clinical or testing experience. The SAICA takes 4 hours of training at a minimum of a master's level to get 95% agreement on co-rated in-person and video ratings. Test-retest and internal consistency data has not been reported. Interrater reliability was a highly acceptable .85 to 1.00 (kappa). The correlation between parent and child report ranged from .10 to .72 on various subscales. Factor analysis of the parent measure results in three factors, accounting for 56% of the variance; the child measure also has three factors, accounting for 51% of the variance. The SAICA has minimally acceptable convergent validity, as it is correlated .52 with the CBCL social competence scale and .64 with the CBCL total problems score. The SAICA problem scale has barely acceptable divergent validity as it correlates .04 to .10 with WISC vocabulary score, .03 to .08 with the PPVT, and .05 to .20 with WISC block design. The SAICA parent version significantly discriminated two different groups from controls. However, the child version failed to discriminate among these groups. The SAICA lacks norms, lacks data on developmental sensitivity, and has not been tested with different ethnic groups. It does have an optional section for use with older children. We do not know whether computer software is available for this measure.

Quality Of Life

Perceived Life Satisfaction Scale (PLSS) (Smith et al., 1987): The PLSS is an adolescent self-report that measure's life satisfaction in 5 areas: (1) *material and physical well-being*, (2) *relationships*, (3) *environment*, (4) *personal development and fulfillment*, and (5) *recreation and entertainment*. The measure consists of 19 items on a 6-point, Likert-type scale. The reading level is unknown, but the fact that the measure was designed and used with adolescents suggests that it is inappropriate for younger children. Using our formula, the PLSS takes 6 minutes to complete. The measure appears to be in the public domain (the items appear in Huebner & Dew, 1993), and except for basic test administration knowledge, the PLSS should not require special training or experience to administer. Test-retest reliability estimates from different samples were barely to somewhat acceptable (.63, .72, and .85), although the retest periods are unknown. Internal consistency was barely acceptable, with a range of .70 to .80 across subscales in one study and .76 to .81 across factors in another study that found a 4-factor solution (Huebner & Dew, 1993). The construct validity of the measure is questionable due to the latter study's disconfirmation of the measure's hypothesized structure and a difficult-to-interpret 4-factor solution that accounted for only 57% of the measure's total variance. The PLSS has a barely acceptable rating for convergent validity as it correlated .58 with the Students' Life Satisfaction Scale. The measure discriminated between "...adolescents in regular classrooms from adolescents referred for mental health services" (Dew & Huebner, 1994, p. 186). Normative, developmental sensitivity, and cultural sensitivity data have not been reported. To our knowledge, computer software has not been produced for this instrument.

Students' Life Satisfaction Scale (SLSS) (Huebner, 1991): The SLSS is a brief, 7-item self-report of global life satisfaction for children and adolescents in grades 3-12. The measure uses a 4-point, Likert-type format. The reading level is unknown, but the measure was designed for use with students in elementary (as early as third grade), middle, and high schools. Based on our formula, the measure should take only 5 minutes to complete. The availability of the measure is unknown, but no special training or experience should be necessary to administer 7 items. Test-retest (1-2 weeks) reliability, in a sample of children in grades 4 through 8, was .74, which is somewhat acceptable. Internal consistency (Cronbach's alpha) ratings for different groups include: (a) .82 in grades 3-7; (b) .86 in grades 8-12; and (c) .85 for black and white students. Item-total correlations have ranged from .49 to .73. These coefficients of internal reliability were considered somewhat acceptable. Factorial evidence of construct validity has been reported: (a) the SLSS reportedly has a one-factor structure that accounted for 55% of the total SLSS variance (Grades 8-12); (b) factor loadings ranged .61 to .83; (c) one factor emerged in independent analyses of black and white students; and (d) all items but one loaded onto the same factor, even when mixed with 20 positive- and negative-affect items. The SLSS correlated (a) .58 with the PLSS; (b) -. 52 with the Nowicki-Strickland Locus of Control Scale-Short Form (greater internal locus of control); (c) .52 with the General-Self subscale of the Self-Description Questionnaire-II (more positive self-image); (d) .57 with the Happiness and Satisfaction subscale of the Piers-Harris Self-Concept Scale; (e) .48 with a one-item parent rating of child's global life satisfaction; (f) .41 and .52 with parent global ratings (black and white students, respectively). Convergent validity was rated barely acceptable because only the highest correlation exceeded the established cut-point of .50. Divergent validity was considered unacceptable due to modest correlations (.29 to .40) with subscales of the Pier-Harris Self-Concept Scale (not theorized to be related). However, in favor of the measure's divergent validity, the SLSS was not strongly related to age (-.09) or grade (-.04). The cultural sensitivity data of the SLSS are mixed. Positive findings include: (a) no gender differences; (b) one factor emerged in separate factor analyses of black and white students; (c) a coefficient of factorial congruence across these groupings was .99; (d) internal consistency was identical (.85) for black and white students; and (e) Pothoff's test for bias revealed no significant difference between concurrent validity coefficients for black and white students. However, other evidence suggests that there are group biases: (a) the SLSS correlated significantly with socioeconomic status (.34); and (b) African-American students reported lower life satisfaction than Caucasian students. Normative data are unavailable to help correct for these potential biases. Developmental data have not been reported. It should be noted that the developer of the instrument recommends using the SLSS only for research purposes until the theoretical base and data base for children's life satisfaction are expanded and better understood (Huebner, 1994). To our knowledge, computer software has not been produced for this measure.

Satisfaction

Satisfaction Scales (Brannan, Sonnichsen, & Heflinger, 1996). The Satisfaction Scales measure satisfaction with children's mental health services within specific service settings and across eight content areas: *access and convenience*, *child's treatment*, *parent services*, *family services*, *relationship with therapist*, *staff responsiveness*, *financial charges*, and *discharge/ transition*. The eight service areas include *outpatient*, *day treatment*, *after-school programs*, *in-home counseling*, *therapeutic home*, *residential treatment center*, and *inpatient*. Separate forms are available for parents and adolescents. Each of these self-reports uses a 5-point, Likert-type format, and consists of 23 to 51 items, depending on the service module. The reading levels for the different forms are unknown, although children are not expected to be capable of completing the measure. The authors report that the scales require up to 10 minutes to complete. Using our formula, we estimate a completion time of 6 to 9 minutes, depending on the module. The scales are available from the authors upon request, and require no special training or experience to administer. Test-retest reliability data have not been gathered. Cronbach's alpha for parent versions ranged .76 to .92 (outpatient), .79 to .94 (inpatient/RTC), .74 to .91 (group home), .77 to .93 (in-home), .57 to .91 (day treatment), .84 to .96 (therapeutic home), .74 to .89 (case management), and .94 to .98 (global satisfaction across service settings). Parent intake assessment alpha coefficients ranged from .77 to .94 across eight content areas. Alphas for adolescent versions (three settings) ranged from .61 to .91 across content areas, and .70 to .88 for the intake assessment questionnaire. We rated these coefficients as unacceptable because there were many low estimates, while internal consistency was barely acceptable for 5 settings, somewhat acceptable for one setting (therapeutic home), and highly acceptable for global satisfaction. Support for content validity is based on the theoretical derivation of the items from the CASSP principles (Stroul & Friedman, 1986). Confirmatory factor analysis supported 4 dimensions of satisfaction (conducted only with the parent version of the outpatient Satisfaction Scale due to the need for a very large sample), with a fit index of .974.

The scales have been tested only with (predominantly white) military families. No significant correlations with race, gender, parent age, education or income were found, except for a small, significant correlation between parent age and access/convenience (older parents more satisfied), and parental gender and child's treatment (fathers less satisfied). Convergent validity was rated unacceptable due to very low, but significant, correlations (.12 to .31) with criteria such as mutually agreed-on termination, parent expectations, and length of treatment. However, criterion-related validity (especially with the chosen variables) may be more difficult to establish than concurrent validity in which scales, that purportedly measure the same constructs, are correlated. Neither normative nor developmental sensitivity data have been reported. Computer software is not available for the Satisfaction Scales.

Self-Esteem

Personal Attribute Inventory for Children (PAIC) (Parish & Taylor, 1978): The PAIC is a 48-item youth self-report measuring perceptions of self. The PAIC was revised by Parish and Rankin (1982) and renamed the Nonsexist Personal Attribute Inventory for Children (NPAIC). The NPAIC format has the subject select 15 items that best describe the self. Information on reading level has not been reported in the literature. However, this is an important issue for a youth-report measure. Using our formula, we estimate that the NPAIC requires 9 minutes to complete. No information was obtained from the literature or catalogues as to the cost of this measure. It is unknown whether training is required to use this measure. One-month test-retest reliability was a barely acceptable .62. A shorter time period would be preferable. Internal consistency was somewhat acceptable at .83 (split-half reliability). The NPAIC has barely acceptable convergent validity, as it is correlated .66 to .67 with the Piers-Harris measure of self-esteem. The NPAIC does not differentiate significantly between normal youth, the physically handicapped, youth with a learning disability, and the emotionally disturbed; however, a teacher-report version yielded significant differences between groups. The PAIC had a normative sample of 1,050 children with separate norms by gender; however, the nonsexist version was introduced after the normative study. There is also a (non-applicable) 1,400-college student normative sample. The NPAIC may be gender-biased and developmentally biased as test-retest reliability varied by grade and gender, with some coefficients below .60. It is unknown whether computer software is available for this measure.

Rosenberg Self-Esteem Scale (RSES) (Rosenberg, 1965): The RSES is a 10-item adolescent self-report, on a 4-point scale that measures self-acceptance. Information on reading level has not been reported in the literature. The RSES requires 3 to 5 minutes to complete. The measure is in the public domain. Training is not required to use this measure. Test-retest reliability has been studied but the time period has not been specified. Test-retest reliability was found to be a barely acceptable .62. Internal consistency (Alpha) was a barely acceptable .73 to .84 across studies. The RSES has highly acceptable convergent validity, as it is correlated .79 with the SDQ-III. It also correlated .72 with Harter's SPPC Global Self Worth. The RSES was not found to discriminate significantly between high-and low-track students. The RSES was normed on 5,024 high school students in 1965; however, we believe it has not been re-normed recently. The RSES does not appear to be developmentally biased (alpha was fairly consistent for middle school and high school students and there were no differences between children in grades 5-8 on scores). No cultural bias has been detected. We do not believe computer software is available for this measure.

Self-Appraisal Inventory (SAI) (Narikawa & Frith, 1972) : The SAI is a 36-item measure, using a yes/no format, read to the child by an examiner and completed by the child. It measures school children's self-concept: *scholastic*, *peer*, *family*, and *general self-concept*. Information on reading level has not been reported in the literature. However, reading level and comprehensibility are important issues for a measure that is read. Using our formula, the SAI requires nine minutes to complete. No information was obtained from the literature or catalogues as to the cost of this measure or any training requirements. Two-week test-retest reliability was an unacceptable .29 to .58 across subscales and .73 on the total score. Internal consistency (KR-20) was an unacceptable .50 to .62 across subscales and .37 for the total scale. Alpha was .41 to .65 across subscales and .82 for the total score. Factor analysis was used to examine the scale's factor structure. The SAI has barely acceptable convergent validity as it is correlated at least .60 with the CDI and the RCMAS. The SAI discriminated significantly between referred and non-referred children. The SAI lacks normative data or data on developmental sensitivity. It appears to be culturally biased as ethnically mixed families displayed lower self-concepts. It is not known whether computer software is available for this measure.

Social Support

Perceived Social Support Scale from Family (PSS-FA) (Procidiano & Heller, 1983); Perceived Social Support Scale from Friends (PSS-FR) (Procidano & Heller, 1983); Perceived Social Support from School Personnel (DuBois, et al., 1992) : The Perceived Social Support Scales measure adolescents' *perceptions of social support* from *family*, *friends*, and *school personnel*. Each scale contains 20 items and uses a yes-no, self-report format. The scales' reading levels are unknown, but they were originally designed for use with adolescents. Based on our formula, each scale requires 6 minutes to complete. The availability of the scales, and the training and experience needed to administer them, are not known. Test-retest reliability data were not found. Internal consistency coefficients have ranged from .85 to .90, which we consider somewhat acceptable. Separate factor analyses of the family and friends scales indicated that only one factor comprises each scale. A variety of variables and measures have been correlated with the scales, affording a confusing picture of convergent and divergent validity. The published correlations include: (a) -. 15 to -. 63 with psychological distress; (b) all three scales .31 to .44 with the Self Appraisal Inventory; (c) all three scales -. 16 to -. 45

with the Children's Manifest Anxiety Scale and Children's Depression Inventory; (d) .15 (family), .26 (friends), and .06 (school personnel) with grade point average; (e) .69, .60, and .75 (family) with identity, self-reliance, and work orientation, respectively; and (f) .37, .26, and .17 (friends) with identity, self-reliance, and work orientation, respectively. The literature is unclear regarding the scales or the criterion-related measures that should be theoretically related or unrelated to the PSS scales. For this reason, we did not code these correlations as evidence of divergent validity, but used our criteria for convergent validity only. Using the PSS-FA and a measure of identity, we rated the convergent validity as barely acceptable. Data on norms, developmental sensitivity, and cultural sensitivity, were not found. To our knowledge, computer software is not available for the PSS scales.

Stressful Events

Daily Hassles Questionnaire (DHQ) (Rowlison & Felner, 1988): The DHQ, patterned after the Daily Hassles Scale, assesses daily life events over the past 2 weeks that could contribute to children's level of stress. The measure is an 81-item, self-report that uses a 4-point, Likert-type scale. The reading level of the DHQ is unknown; but the authors report that it can be used with adolescents and "school-age" children. We estimate the measure requires 12 minutes to complete. The training, and experience needed to administer the measure are not known. Test-retest reliability data were not found, but the conventional 2-week test-retest period may be inappropriate for the DHQ since its construct may change frequently. (Testing reliability using alternate forms may be more appropriate in this case.) Cronbach's alpha was highly acceptable at .95. The measure was developed after obtaining open-ended responses from school-age children. Although 44 items were rated by masters - and doctoral - level clinical psychologists as overlapping with signs of emotional distress, the remaining situations described in the measure may or may not represent events that typically introduce stress into children's lives. The DHQ correlated -.45 with the Self-Appraisal Inventory, .48 with the Children's Manifest Anxiety Scale, .44 with a measure of negative life events, and .50 with the Children's Depression Inventory. Correlations with the Life Events Checklist, a measure of psychological distress and family support ranged .28 to .33, .33 to .58, and -.28 to -.38, respectively. Convergent validity received a rating of "unacceptable" due to the frequent correlations below .50. Divergent validity was rated minimally acceptable as the DHQ correlated -.02 to -.24 with grade point average. Normative data and data on developmental and cultural sensitivity, were not found. To our knowledge, computer software is not available.

Hassles Scale for Children (HSC) (Parfenoff & Jose, 1989). The 49-item HSC measures the type and degree of stress in children's daily lives using a 3-point, Likert-type scale. The reading level of the measure is unknown, although the authors' intent was to use the scale with pre-adolescent children. Completion time is estimated at 9 minutes. The availability, training, and experience needed are not known; however, basic training and experience in test administration should be adequate. Two-week test-retest reliability (.74) was deemed somewhat acceptable. Internal consistency (.88) was considered mostly acceptable. Validity data were not found, nor were normative, developmental or cultural sensitivity data. To our knowledge, computer software is not available.

Therapeutic Process

Family Involvement Items (Baker, Blacher, & Pfeiffer, 1993). These are 4 items, not a named measure, completed by a residential facility staff member who knows the child and family. The items measure 4 dimensions of family involvement and provide an overall index (sum total): *participation in social and educational activities at facility* (on a 3-point scale), *frequency of phone calls to staff or child, frequency of visits to child, and frequency of visits home or overnight with family* (last three domains on a 6-point scale). Information on reading level is not applicable for these rated items. Using our formula, it is estimated that the items require 5 minutes to complete. As far as we can tell, the items appear to be in the public domain. It is unknown whether training is required to use this measure. There are no test-retest or interrater reliability data. Internal consistency (alpha) was a somewhat acceptable .87. The items have minimally acceptable convergent validity as they are correlated -.57 with family residence distance from the residential facility. The items were found to significantly discriminate families with children with dual diagnoses. The items have not been normed and there are no data on developmental sensitivity. The items do appear to be culturally biased as they were correlated -.50 with SES and .34 with race. Computer software is not available for these items.

Treatment Acceptability Questionnaire (TAQ) (Hunsley, 1992). The TAQ is an adult or adolescent self-report that assesses the treatment acceptability of psychological treatment using 6 items on a 7-point scale. The measure can be used across treatment modalities, such as individual and group therapy. The reading level, availability, training, and experience needed to administer the measure are unknown. The developer of the instrument used a convenience sample of psychology undergraduate students to rate a college student case vignette and another sample to rate a child case vignette. The 2-week test-retest coefficient was .78 among students who rated the college student case vignette. Test-retest reliability was rated as somewhat acceptable, but it should be noted that this estimate is not based on an evaluation of children's treatment. Internal consistency was .74 and .80 across administrations (evaluating the college student vignette), and .81 evaluating the child case vignette. Rating the internal reliability of the child vignette ratings only, internal consistency was somewhat acceptable. Items were generated based on a theory of treatment acceptability reported in the literature. Convergent validity was rated highly acceptable as the TAQ correlated .87 with the TEI (see below). The measure also correlated .59 with the Evaluative (good-bad) subscale using the Semantic Differential method. Divergent validity would be considered highly acceptable as the TAQ correlated only -.02 and -.06 with *Potency* (strong-weak) and *Activity* (active-passive) subscales using the Semantic Differential method. Hunsley expected a significant relationship between the TAQ and these dimensions and did not consider that they may be divergent constructs. Data on norms, developmental sensitivity, and cultural sensitivity were not found. To our knowledge, computer software is not available for this measure.

Treatment Evaluation Inventory (TEI) (Kazdin, 1980) and **TEI-Short Form** (TEI-SF) (Kelley, Heffer, Gresham, & Elliot, 1989): The TEI is a parent self-report of how acceptable, likable, suitable, cruel, unfair, and effective their child's mental health treatment is. Two forms are available: a 15-item version on a 7-point scale, and a simplified 9-item version on a 5-point scale (TEI-SF). The reading level, availability, training, and experience needed to administer the measure are unknown. Average completion times are 3.5 minutes for the TEI and 2.2 minutes for the TEI-SF. Reading grade levels for the TEI and TEI-SF are 5.1 and 4.2, respectively. The TEI-SF is in the public domain, but the availability of the TEI is not known. Except for basic test administration knowledge, special training or experience should not be required to administer either the TEI or the TEI-SF. Test-retest data were not found. Internal consistency was .89 on the TEI and .85 on the TEI-SF, both somewhat acceptable. The median inter-item correlation on the TEI was .67. Factor analysis was used to design the 15-item TEI, but a replication study (Kelley et al., 1989) yielded two factors instead of the original uni-dimensional structure reported by Kazdin. To estimate the similarity of factor patterns across parent and student raters, coefficients of congruence were computed (.95 and .87 for the two factors). The high congruence between these raters suggests that the test has a consistent factor structure. The correlation of the TEI with the TAQ (see above) indicates highly acceptable convergence (.87). The TEI has been shown to differentiate child treatments and child management techniques significantly. However, the two versions have yielded unequal evaluation scores for some types of treatment. In one study, 71% of subjects liked the TEI-SF more than the TEI. Normative and developmental sensitivity data were not found. Scores differed as a function of parent income and race, suggesting cultural differences on ratings. Moreover, mothers and fathers rated the acceptability of (5 of 6) child management techniques differently. To our knowledge, computer software is not available for either the TEI or TEI-SF.

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

Mental health services differ in their resources for, and receptiveness to, outcome measurement. For that reason, we propose a modular strategy. Services wishing to collaborate in the development of outcome measurement can choose to adopt one or more of the following modules:

The Baseline-Followup Module (BFM)

The Background Module (BM)

The Concurrent Module (CM)

These modules serve different purposes. The Baseline-Followup Module (BFM) primarily provides data that allows the evaluation of service effectiveness for the purpose of administrative and policy decision making. The Background Module (BM) provides information to clinicians concerning those moderating factors that influence the degree to which treatment will be effective in the individual case. In addition, this module can provide useful information about the characteristics of the consumers served. The Concurrent Module (CM) provides feedback to clinicians concerning the progressive effectiveness of individual treatment. Both the BM and CM provide important information that can link clinical processes to outcomes and assist in individual case management. Each of the three modules addresses a different aspect of quality improvement.

Borrow, Buy or Build?

The Baseline-Followup Module

Table V.1 provides information related to the decision whether to borrow, buy or build the Baseline-Followup Module (BFM). The criteria used to make this decision were as follows: (1) the instrument should take no longer than 30 minutes to complete; (2) the module should contain the domains selected, namely *functional impairment*, *acuity*, *parent-child relationship*, *symptom severity*, *quality of life*, and *satisfaction*; (3) the instrument or instruments that comprise the BFM should be on the “best instruments” list; and (4) the instrument should be available at no cost. We have compared and contrasted five possible modules. The first column, labeled BFM1, presents a module composed of several instruments we have already reviewed. This module would require both buying and borrowing: the Family Environment Scale would need to be purchased but the rest of the instruments are in the public domain. We do not recommend BFM1 because of its cost and because it exceeds the 30-minute time limit. BFM2 uses the ATOM as the core instrument. It would require the addition of measures in three domains, and would also exceed the time limit. BFM3 far exceeds the time limit and lacks measures of *parent-child relationship, quality of life,* and *satisfaction*. BFM4 is based on another comprehensive instrument, the SAICA. BFM4 would have to be supplemented in several areas, and exceeds the time limit. BFM5 has the Ohio Youth Problems, Functioning and Satisfaction Scales as its core instrument and would require the addition of an acuity measure, and refinement of the *quality of life* and *quality of the parent-child relationship* domains. However, BFM5 is the only option that meets the important time criterion (no longer than 30 minutes to complete).

Because BFM5 best meets the aforementioned criteria for a baseline-followup module, we recommend that the Ohio Youth Problems, Functioning and Satisfaction Scales serve as core of the BFM. We further recommend that the Family APGAR and the Students' Life Satisfaction Scale be added to refine and supplement the Ohio Scales, and that an acuity measure be borrowed or built. Such a composite module could not be adopted without further psychometric developmental work. Extensive work would be needed to assure that the instrument is reliable, valid, and culturally appropriate in Australia. In addition, the instrument would have to be normed in Australia for maximal usefulness. Essentially, borrowing instruments provides a pool of items that have been shown to be valid in other situations.

Table V.1.—Baseline Follow-up Modules

Domain	BFM1	Minutes	BFM2	Minutes	BFM3	Minutes	BFM4	Minutes	BFM5	Minutes
Multidimensional	None	-	Adolescent Treatment Outcomes Module	28	Behaviour Assessment System for Children	30; see app	Social Adjustment Inventory for Children and Adolescents	30	Ohio Youth Problems, Functioning, and Satisfaction Scales	10
Functional Impairment	Columbia Impairment Scale	5	Y*	-	Y*	-	Y*	-	Y*	-
Acuity	Acuity measure to be developed	*2-3*	Acuity measure to be developed	*2-3*	Acuity measure to be developed	*2-3*	Acuity measure to be developed	*2-3*	Acuity measure to be developed	*2-3*
Quality of the Parent-Child Relationship	Family Environment Scale	10	The ATOM measures family burden, not the quality of the parent-child relationship	10	The BASC does not measure the quality of the parent-child relationship	10	The SAICA does not measure the quality of the parent-child relationship	10	The Ohio Scales minimally measure the quality of the parent-child relationship	10
Symptom Severity	Paediatric Symptom Checklist	5; parent	Y	-	Y	-	The SAICA does not measure symptom severity	5	Y	-
Quality of Life	Students' Life Satisfaction Scale	*5*	The ATOM does not measure quality of life	*5*	The BASC does not measure quality of life	*5*	The SAICA does not measure quality of life	*5*	Y (The Ohio Scales measure well-being and hopefulness)	0
Satisfaction	Satisfaction Scales	*10*; see app	Y	-	The BASC does not measure satisfaction	*10*	The SAICA does not measure satisfaction	*10*	Y	-
All	CIS, acuity measure, FES, PSC, SLSS, and Satscales	*37-38*	ATOM, acuity measure, FES, and SLSS	*45-46*	BASC, acuity measure, FES, SLSS, and Satscales	*57-58*	SAICA, acuity measure, FES, PSC, SLSS, and Satscales	*62-63*	Ohio Scales, acuity measure, Family APGAR, and SLSS	*22-23*

Note: "Y" indicates that the domain is measured by a subscale of the multidimensional measure within each respective module, thereby averting the need for additional measures for that domain. An "*" indicates that the multidimensional measure represented measures functional competence in addition to functional impairment. Italicised items are estimates.

The Concurrent Module.

We have already identified four areas for concurrent measurement: (1) *treatment goals*; (2) *the strategy and tactics of treatment*; (3) *symptom severity* and *functional impairment*; and (4) *therapeutic alliance*, *readiness for change*, and *adherence to treatment*.

As shown in Table V.2, none of the “best instruments” we reviewed assesses goals the way we have recommended. We have not found any instrument that satisfactorily describes *the strategy and tactics of treatment*. We recommend the design de novo of instruments that measure *goal attainment* and describe *the strategy and tactics of treatment*. The Ohio Youth Problems, Functioning and Satisfaction Scales can be used to measure progress in *symptom severity* and *functional impairment* concurrently with treatment. While it has not been tested for short time intervals and repeated measures, its brief administration time makes the Ohio Scales an excellent candidate for further development. The Ohio Scales also has forms for youth, parent, and therapist. We have not been able to identify satisfactory instruments that measure *therapeutic alliance*, *readiness for change*, and *adherence to treatment*. We recommend the design de novo of instruments for these constructs.

The Background Module

Table V.3 shows possible configurations for a Background Module. Completed by the carer this module describes the child’s background. Background information is important not only for the descriptive information it offers, but also because it may be helpful in determining the factors that moderate treatment success. Although these factors might not be addressed directly in treatment, they could affect the outcome of treatment. Unlike the variables in the BFM (which should change due to treatment), background domains may or may not change as a result of treatment. However, if the family has poor resources or if the child has been mistreated, the level of outcomes could be affected. No combination of existing “best instruments” would be adequate for a background module. We recommend that the background module be designed de novo.

Table V2.—Concurrent Modules

Domain	CM1	Minutes	CM2	Minutes
Multidimensional	None	-	Ohio Youth Problems, Functioning, and Satisfaction Scales	10
Therapeutic Relationship According to Children	Treatment Acceptability Questionnaire	5	The Ohio Scales do not measure the therapeutic relationship according to children	5
Therapeutic Relationship According to Parents	Treatment Evaluation Inventory	4	The Ohio Scales do not measure the therapeutic relationship according to parents	4
Functional Impairment	Columbia Impairment Scale (CIS)	5	Y	-
Symptom Severity	Paediatric Symptom Checklist (PSC; parent report)	5	Y	-
Treatment Adherence	Treatment adherence measure to be developed	To be determined	The Ohio Scales do not measure treatment adherence	To be determined
Treatment Goals	Treatment goals measure to be developed	To be determined	The Ohio Scales do not measure treatment goals	To be determined
Tactics of Treatment	Treatment tactics measure to be developed	To be determined	The Ohio Scales do not measure treatment tactics	To be determined
All	TAQ, TEI, CIS, PSC, and measures of adherence, treatment goals, and treatment tactics	10 minutes for child, 9 minutes for parent, plus time to complete treatment adherence, goals, and tactics measures	Ohio Scales, TAQ, TEI, and measures of adherence, treatment goals, and treatment description	15 minutes for child, 14 minutes for parent, plus time to complete treatment adherence, goals, and tactics measures

Note: "Y" indicates that the domain is measured by a subscale of the Ohio Scales, thereby allaying the need for a supplementary measure for the domain represented within the row.

Table V3—Background Modules

Domain	BM1	Minutes	BM2	Minutes	BM3	Minutes	BM4	Minutes	BM5	Minutes
Multidimensional	None	-	Adolescent Treatment Outcomes Module	28	Behaviour Assessment System for Children	30; see app	Ohio Youth Problems, Functioning, and Satisfaction Scales	10	Social Adjustment Inventory for Children and Adolescents	30
Safety of Environment	Safety measure to be developed	5	Safety measure to be developed	5	Safety measure to be developed	5	Safety measure to be developed	5	Safety measure to be developed	5
Stressful Events	Hassles Scale for Children	9	The ATOM does not measure stressful events	9	The BASC does not measure stressful events	9	The Ohio Scales do not measure stressful events	9	The SAICA does not measure stressful events	9
Family Resources	Family Resource Scale	7	The ATOM does not measure family resources	7	The BASC does not measure family resources	7	The Ohio Scales do not measure family resources	7	The SAICA does not measure family resources	7
Maltreatment	Child Abuse Potential Inventory	20	Child Abuse Potential Inventory	20	Child Abuse Potential Inventory	20	Child Abuse Potential Inventory	20	Child Abuse Potential Inventory	20
Other Pertinent Client Background Characteristics	Domain(s) and measure(s) to be determined	3-5	Domain(s) and measure(s) to be determined	3-5	Domain(s) and measure(s) to be determined	3-5	Domain(s) and measure(s) to be determined	3-5	Domain(s) and measure(s) to be determined	3-5
All	Safety measure, HSC, FRS, CAPI, and other measures	44-46	ATOM, safety measure, HSC, FRS, CAPI, and other measures	72-74	BASC, safety measure, HSC, FRS, CAPI, and other measures	74-76	Ohio Scales, safety measure, HSC, FRS, CAPI, and other measures	54-56	SAICA, safety measure, HSC, FRS, CAPI, and other measures	74-76

Building the Concurrent Module (CM)

Since we recommend borrowing scales for the development of the BFM, we do not provide details. However, concurrent measurement is a very new concept and we recommend that there should be more development in this area. Thus, we provide more information about its development.

Goals

Goal Attainment Scaling

Kiresuk and Sherman (1968) introduced Goal Attainment Scaling GAS (1968) as a method of outcome assessment, based on the contention that there are no universal definitions of human service goals. Rather, Kiresuk and Sherman suggested, service needs are best defined according to the unique problems of each client at the time of treatment. GAS measures incremental progress toward goals. It is a flexible, idiographic technique, tailored to individual needs, and designed for the evaluation of program effectiveness on the basis of the extent to which individual client goals, established at intake, have been achieved at a specified later date. Kiresuk, Smith and Cardillo (1994) have recently updated the history, implementation, applications, and psychometric qualities of GAS. The following methodology is applied in GAS (Kiresuk and Sherman (1968).

1. An independent goal selector sets treatment goals in accordance with the client's problems and needs.
2. The goals are weighted to reflect their importance.
3. Observable and measurable goal attainment indicators are selected.
4. The follow-up time is designated.
5. Expected levels of outcome are identified and rated at zero.
6. Four additional levels of outcome are identified, ranging from "much less than expected" (-2) to "much better than expected" (+2), producing an exhaustive and internally consistent continuum of all possible outcomes.
7. A Goal Attainment Follow-Up Guide (GAFG) is completed and reviewed to eliminate vagueness, incompleteness, overlap between goals, and multi-dimensional scaling.

8. The GAFG is used to assess the degree of goal attainment.

9. Average goal attainment is calculated in accordance with a mathematical formula.

Goal-Directed Treatment Planning

Goal-Directed Treatment Planning was introduced by Nurcombe and Gallagher (1986) and subsequently elaborated in a number of papers (Nurcombe, 1987(a), 1987(b), 1987(c), 1989). The technique of goal-direction was originally designed in order to facilitate efficient treatment planning during brief hospitalization; however, with suitable modification, it is applicable to any level of patient care.

In order to clarify the concepts behind Goal-Directed Treatment Planning, the following definitions are required:

- A *diagnostic formulation* is a summary statement that integrates the following issues: biopsychosocial predisposition; precipitation; the current pattern of biopsychosocial disequilibrium as reflected in symptom pattern and coping style; the factors that perpetuate the current disequilibrium; prognosis; and the patient's and family's strengths, resources and potentials.

- A *pivotal problem* is a physical disorder, symptomatic behaviour, dysfunction, unresolved psychological issue, functional impairment, or perpetuating factor that if altered in the desired direction, would undercut one or more of the vicious circles that cause the patient's presenting disequilibrium. The clinician chooses those pivotal problems that, if altered, are likely to produce the greatest benefit. Pivotal problems are restated as goals.

- A *goal* represents what the clinician aims to achieve in collaboration with the patient. Goals may be related to stabilization (e.g., "Eliminate suicidal ideation"), reconstruction (e.g., "Resolve conflict concerning past sexual abuse"), remediation (e.g., "Improve reading skills"), or compensation (e.g., "Promote artistic talent").

- An *objective* anchors a goal by stating it in concrete, observable, and (preferably) measurable terms. Thus, the goal "Reduce the intensity of depressive mood" could be objectified in terms of an improvement of mood on mental status examination, or a reduction of the depression score on an appropriate self-administered questionnaire or parent-report measure.

- A *treatment modality* is the kind of treatment provided (e.g., medication, systemic family therapy, cognitive behaviour therapy).
- Within each treatment modality the clinician implements a *treatment strategy*, that is, a plan of intervention (e.g., within the modality of cognitive behaviour therapy, the clinician might apply relaxation, exposure and response prevention). From minute to minute the clinician implements *treatment tactics* (e.g., microcounselling skills, focussing, or interpretation).
- For each goal, the clinician estimates the time required for goal attainment, expressed as a *target date*.
- Based on the stated objectives, a *treatment outcome monitoring system* is established, in order to check at regular points the degree to which the goals are being approximated, and to provide progressive information to the clinician regarding the fulfillment or otherwise of the treatment plan. Treatment ceases when the goals are achieved.
- *Negotiation* is a discussion between patient, parents and clinician whereby the diagnostic formulation is agreed upon, and the goal-directed treatment plan is contracted. Goals, objectives, treatment modalities, potential risks, potential benefits, alternative treatments, target dates and costs are shared, and the patient's and family's goals are incorporated or embodied in the treatment plan. Negotiation is the process whereby informed consent is obtained and a therapeutic alliance established.
- *Implementation* is the process whereby the treatment plan is effected.

The operational steps of Goal-Directed Treatment Planning are as follows:

1. Formulate a biopsychosocial diagnosis.
2. Extract pivotal problems from the formulation.
3. Restate pivotal problems as treatment goals.

4. For each goal, stipulate one or more objectives.

5. Select the therapy or therapies most appropriate to each goal.

6. Designate a target date for each goal.

7. Select methods of evaluation of each goal in accordance with the objectives.

8. Negotiate with the patient and family, the diagnostic formulation and treatment plan.

9. Implement the treatment plan.

10. Monitor treatment progress regularly by means of the evaluation system.

11. Revise the treatment plan if new information demands a change in the diagnostic formulation, or if progress stalls or is reversed.

12. Terminate treatment when the stipulated goals/objectives are attained.

Goal-Directed Treatment Planning was introduced with the aim of replacing treatment-oriented planning (an intuitive mode of decision-making lacking clarity and accountability), and problem-oriented treatment planning (Weed, 1968) which, in mental health, tended to deteriorate into lists of fragmentary, superficial “behaviours” (Nurcombe, 1989). Goal-direction has the advantage of clarity of purpose and communication (particularly helpful for multidisciplinary teams). It promotes accountability, efficiency, and outcome research. However, it is not the “natural” method of planning. Naturalistic, intuitive planning matches treatment to problem without specification of goals and objectives. Naturalistic planning monitors treatment progress in accordance with a global impression of change rather than a defined evaluation system, creating a risk of therapeutic “drift”.

Goal-direction has not been easy to introduce. Educational workshops are required. As with any innovation, the adoption of the method is impeded by conservatism and entropy. However, goal-direction is sufficiently flexible to accommodate different theoretical approaches; indeed, it has the potential to promote a practical eclecticism.

What is the relevance of Goal-Directed Treatment Planning for outcome measurement? As with Goal Attainment Scaling, Goal-Directed Treatment Planning emphasises the importance of individualised goals, lends itself to clinician-patient goal contracting, and encompasses the possibility of regular concurrent measurement in relation to treatment objectives. We believe that, for the following two reasons Goal-Directed Treatment Planning is potentially superior to Goal Attainment Scaling as process/outcome measure:

1. The goals of GAS are formulated as behaviours, and as such are often superficial, missing the deeper issues that commonly underlie symptoms or impairment. (For example, *running away from home* could be the manifestation of *unresolved conflict concerning sexual abuse*, or of a *mood disorder*, or of the child's *response to domestic violence*.) In contrast to the goals of GAS, the goals of Goal-Directed Treatment Planning are derived from a diagnostic formulation that combines symptoms, coping patterns, precipitating factors, and perpetuating factors. The goals of Goal-Directed Treatment Planning are expressed in broad, abstract terms but anchored by behavioural objectives.

2. Goal Attainment Scaling requires the subjective rating on a 5-point scale (-2, -1, 0, +1, +2) of the degree to which each expected goal was attained. Both goal setting and goal attainment scaling are subjective and of uncertain validity. Goal-Directed Treatment Planning involves the objectification of goals in a measurable form. Goal-Directed Treatment Planning lends itself to the assessment of goal attainment in a potentially more reliable and valid form.

It would be possible to provide clinicians with a menu of goals from which they could select the five or six most relevant to the pivotal problems derived from the diagnostic formulation. For each goal, an objective measurement device could be designed. For example, for obsessive-compulsive thought disorder, the Y-BOCS (Goodman, Rasmussen, Riddle & Rapoport, 1991), a semistructured interview, could be used as a progressive, objective, monitoring instrument. The ultimate objective, for example, would be to reduce obsessive-compulsive thinking to below a stipulated level on the Y-BOCS interview.

It would take time to design, pilot and implement a menu of goals each of which has a specified, objective measurement instrument. No system currently available embodies the diversity and comprehensiveness demanded by contemporary mental health services. We believe that concurrent outcome measurement is sufficiently important to warrant the design and development of specific measures for the goals of treatment. We recommend that goal-directed treatment planning be monitored by providing clinicians with a comprehensive menu of goals for each of which a specific measure is designed. Tracked at regular intervals, goal attainment will be estimated by the degree to which the particular behaviour or symptom score objectifying each goal approximates the expected level of improvement. In order to select a behavioural measure for each goal, extensive pilot testing of borrowed, purchased, or freshly designed instruments will be required. Particular attention must be paid to each instrument's sensitivity to change.

Goal-directed measures assess both process and outcome. The use of goals allows clinician, patient and family to negotiate a clear treatment contract and to estimate whether or not treatment is working. The use of standardized goal-directed measures is a potential advance on Goal Attainment Scaling (Kiresuk & Sherman, 1968) since the goals set according to GAS were simplistically behavioural, subjectively scored, and psychometrically unsound.

The Modality, Strategy, Tactics, and Dosage of Treatment

As noted earlier, without a clear description of the treatment provided it is difficult to improve treatment even with the best outcome measurement. Since few clinicians use manuals or protocols in their practice, it is necessary to develop an approach to describing treatment so that it is possible to relate treatment to proximal and distal outcomes. It is important not only that the modality of treatment be recorded, but also that, within that modality, the strategy, tactics, timing, and dosage of treatment be specified. The *modality* of treatment refers to the broad category of treatment type (e.g., psychoactive medication, conjoint family therapy, cognitive behaviour therapy, or psychodynamically oriented individual psychotherapy). The *strategy* of treatment refers to the particular mode or modes of treatment within each modality (e.g., within behaviour therapy, systematic desensitisation). The *tactics* of treatment refers to those interventions undertaken minute to minute during each therapeutic session (e.g., clarificatory interpretation). The *timing* and *dosage* of the therapy refer to the frequency of therapeutic sessions (e.g., individual psychotherapy, twice per week) and, if appropriate, the quantity of the therapeutic agent.

There is no extant instrument in this area. The development of a new instrument will have to proceed from a conceptual model of the work of a therapist. A simple system could involve a concise list that the clinician can review after each treatment session. In order to make the task as convenient as possible, the clinician should have available an exhaustive “menu” of treatment modalities, strategies and tactics, the relevant items of which could be checked after each treatment session.

Therapeutic Alliance, Treatment Adherence, and Motivation to Change.

Therapeutic Alliance

Therapeutic alliance has to do with the strength of the therapeutic relationship between a client and a therapist (Horvath, 1994). The therapeutic alliance has both affective and cognitive components (Bordin, 1979), including the client’s perceptions of the therapist (as helpful) and of therapy itself (as teamwork) (Luborsky, 1994). We have identified nine dimensions of the therapeutic alliance, as follows:

1. Bond felt with therapist: The degree to which mutual attachment, trust, respect, liking, commitment, and acceptance are felt in the relationship (Bordin, 1979; Horvath & Luborsky, 1993; Klee, Abeles, & Muller, 1990).

2. Perception of supportiveness/helpfulness of therapist: The degree to which the client perceives the therapist as providing or capable of providing the help that is needed (Luborsky, 1994).

3. Affect/attitude toward therapeutic tasks: The degree to which both client and therapist perceive that the work of therapy is potentially helpful (Bordin, 1979).

4. Evaluation of therapist perceptiveness: The client’s perception of how accurately the therapist is able to discern the client’s thoughts and emotions.

5. **Evaluation of clarity of therapy / therapist:** The degree to which the client perceives the therapist makes therapy understandable.
6. **Evaluation of therapist attitude (positive and negative):** The client's perception of whether the therapist has a positive or negative attitude, judged by what the therapist says or does in therapy (Najavits & Strupp, 1994).
7. **Openness toward therapist:** The degree to which the client feels safe enough in the therapeutic relationship to reveal private matters.
8. **Working collaboration between therapist and client:** The client's sense that therapy is a collaborative process (Luborsky, 1994).
9. **Agreement on goals:** The client's perception that he and the therapist agree on treatment goals and their importance (Bordin, 1979).

Using this conceptual framework we believe a brief and valid instrument should be developed that measures therapeutic alliance within the concurrent module.

Treatment Adherence

Treatment adherence involves the client's willingness to participate and cooperate in treatment and his or her preparedness to undertake therapeutic work ("homework") between therapy sessions. Treatment adherence is a composite of *willingness/resistance* (to attend), *cooperativeness/uncooperativeness* (during sessions), and *willingness/unwillingness to undertake homework* (between sessions). Though conceptually related to motivation (discussed in the next section), adherence should be distinguished from it. For instance, poor *participation* involves specific behaviours, such as arguing with parents about attendance, or refusing to accompany the parent to sessions. *Motivation* denotes the clients' feelings about, and desire to resolve, a personal problem. Treatment adherence should be associated with positive motivation and alliance and favourable outcome.

Using this conceptual framework we believe a brief and valid module should be developed that incorporates a measure of treatment adherence concurrently with treatment.

Motivation to Change

Motivated clients perceive their participation in treatment as voluntary. Research indicates that motivated, voluntary clients progress faster, remain in treatment longer, and improve more than those who feel coerced into therapy (Prochaska, DiClemente, & Norcross, 1992). Prochaska and DiClemente (1986) posited five motivational stages that describe progress toward problem solution : *precontemplation* (not willing to admit a problem and not ready to change); *contemplation* (willing to admit a problem but not ready to change); *preparation* (willing to admit a problem and preparing for active change); *action* (actively working on the problem); and *maintenance* (sustaining change).

We conceive of two measures of motivation to be included in this module: an instrument that assigns clients to one of the five stages of change hypothesized by Prochaska and DiClemente (1986), and an instrument designed to measure progressive change, similar to the University of Rhode Island Change Assessment (URICA; McConnaughy, Prochaska, & Velicer, 1983; McConnaughy, DiClemente, Prochaska, & Velicer, 1989). The URICA asks respondents to rate on a five-point Likert scale their agreement or disagreement with statements about changing a particular problem behavior. The URICA can be scored to assess the client’s endorsement of four of the five above stages of change (excluding *preparation*) and the strength of their motivation to change. The URICA was originally developed for adults in outpatient therapy, but it has been used effectively with adolescents in treatment for alcohol abuse (Migneault, Pallonen, & Velicer, 1997). Mean internal consistencies for each subscale are somewhat acceptable (.88 in McConnaughy et al., 1983; .82 in McConnaughy et al., 1989). Each item on a 24-item short version loaded satisfactorily on the appropriate stage-of-change subscale (Carbonari et al., 1994).

These measures, though promising, would not be usable in their current form. They have not been tested concurrently with treatment, nor do data support their use with the full range of adolescent mental health problems. Further instrument development is required.

General Symptomatology and Functional Impairment

As a general indicator of change, a composite measure of *general symptoms* and *functional adaptation* is required. This measure can be taken or adapted from the baseline-followup module. We recommend that the Ohio Youth Problems, Functioning and Satisfaction Scales, described earlier, be modified for concurrent (e.g., fortnightly) data collection.

Summary

We recommend that a concurrent measure be designed and tested. This measure would be a composite of the following: *goals of treatment; the modality, strategy, tactics, timing and dosage of treatment; therapeutic alliance, motivation to change, treatment adherence; and general symptomatology and functional impairment*. The concurrent measure would be completed by both parent and adolescent or by parent alone for children younger than 12 years of age.

The appropriate frequency of concurrent measurement is unclear, since it is not known how much change could be detected during outpatient treatment on (for example) a weekly, fortnightly, or monthly basis. The design and testing of the composite concurrent measure should be undertaken in clinics or services that are receptive to the idea of concurrent measurement and with clinicians who would regard instrument design and piloting as a collaborative project.

Building the Background Module (BM)

Consistent with the criteria for the BFM, the background module should be easy to administer and take less than 30 minutes for each consumer to complete. This module does not require the involvement of the clinician. We have identified the following as the core background domains:

- Safety of the environment
- Stressful events
- Family resources
- Maltreatment
- Other background characteristics

The best way to build this instrument is to survey service settings to determine the background information they currently collect, the information they find most and least useful, and the new information they would like to obtain.

Political Issues

Clinical Perspectives

The mounting costs of health care have prompted the current interest in service evaluation and outcome measurement. Outcome accountability and health services evaluation are being introduced at a time when the political philosophy of economic rationalism is dominant, with its emphasis on value for money, the application of economic restraints, and the promotion of competition in the service of efficiency.

Health services evaluation is accompanied by a language borrowed from the world of commerce indicating that health care is a commodity to be bought, sold, and exposed to the forces of the free market. Thus, social workers, psychologists, psychiatrists and other clinicians become “providers” (who are presumably interchangeable), “patients” or “clients” become “consumers” or “customers”, the medical connotation of “treatment” is replaced by the industrial implications of “management”, and “medicine” is replaced by the demystified term “health care (industry)”. “Health care” is portrayed as an industrial process akin to a factory assembly line, with inputs, outputs, operatives, processes, and quality assurance.

Clinicians resist these industrial terms, perceiving them as imposed by outsiders who are unaware of, or unsympathetic to, the values of psychiatrists, psychologists, social workers, psychiatric nurses and other clinicians. Moreover, the new language is characteristically imbedded in blizzards of paper aimed to impose radical changes about which frontline clinicians have not been consulted.

These political considerations have serious implications for the introduction of an outcome measurement system, particularly in regard to clinician acceptance. We will discuss later how the typical scepticism of clinicians might be acknowledged, addressed, and, in some cases, reversed.

Administrative Perspectives

The resistance of clinicians to outcome accountability is counterbalanced by another point of view. Clinicians are seen to guard their autonomy closely. They ask administrators and policy makers to base resource allocation on the service needs that they, the clinicians, consider essential. Consequently, clinicians may have little credibility with policy-makers. Consumers have more credibility, but they lack technical knowledge and are often unaware of how to be heard. Seeking rational criteria for decision-making, mental health administrators have turned to industrial systems management theory. The patient ("consumer") is the input. Diagnosis and treatment are the manufacturing process. The clinician ("provider") is the assembly line worker. The consumer, hopefully better at the end of treatment, is the product. Efficient production requires the application of reliable, standardized production techniques. Product quality is improved by monitoring the production process, and by feeding information back to managers and operatives in such a way that faulty or inefficient procedures can be eradicated or improved.

There are two problems with the industrial analogy. Patients are not products. Unlike widgets (which have no concern for those who assemble them) patients develop relationships with their therapists, relationships that are inseparable from, indeed integral to, the treatment process. Clinicians are deeply protective of this therapeutic connection. Concurrent outcome monitoring could be perceived as intruding upon the therapeutic relationship, threatening the trust that is the keystone of the treatment alliance.

Consumer Perspectives

The third party to this evolving process is the consumer. Consumers and carers have a vital interest in the quality of services. Are the processes of intake, diagnosis, and therapy respectful, individualized, efficient, and effective? The carer is likely to be impressed by the accessibility and availability (or otherwise) of services provided for the child or adolescent. Recently, consumers have begun to form organisations that pressure government to supply more effective and responsive services; however, the parents of children with mental illness have been slower than adult consumers to unite in this manner.

A Potential Resolution

How can the interests and concerns of these three groups be brought together? Administrators look for assistance in deciding how to allocate resources, particularly when the expansion of services is requested or innovative services proposed. How much will it cost? What is the evidence that it works? How well does it work, for whom, and in whose hands? Are there alternative, equally effective, more efficient, less expensive treatments? Is the current workforce able to deliver the new service? If not, what education is required? These are urgent questions. In time, concurrent and baseline-followup measurement will assist rational administrative decision making.

The consumer wants services that are available and accessible, developmentally and culturally sensitive, efficient and effective. Like clinicians, consumers suspect that administrators aim to slash mental health services in favour of other programs that are politically more attractive, more vociferous, or more prestigious. They are cautious about any infringement on their confidentiality unless it has a worthwhile purpose.

The clinician is aware that existing mental health services for children and adolescents are inadequate. Services need to expand and subspecialise, offering a variety of proven and effective treatment programs in different levels of a continuum of care. The expansion required exceeds the financial and human resources currently available. It would be irresponsible for administrators to embark on change without the means of checking whether the expansion or addition of services is effective.

Conceivably, concurrent and baseline and follow-up measurement is the key to change. Each of the chief stakeholders has or could have an interest in a measurement system: the administrator to get value for money and provide the best services possible; the consumer to assist in the improvement of services; and the clinician to know whether treatment is working or has worked. What is needed is an educational process involving all stakeholders in a meeting of minds that will ensure that the enterprise begins and is sustained. Before embarking on the educational venture, however, each of the three main stakeholders will need to agree to relinquish a part of his autonomy: the consumer would have to agree to being measured; and the clinician and administrator would have to accept their joint and separate accountability for the effectiveness of services.

Educating the Stakeholders

As Hernandez and Hodges (1996) suggest, the following prerequisites apply to the introduction of an innovative system like concurrent and baseline and follow-up measurement.

- Sustained, energetic, committed leadership.
- A receptive political climate among stakeholders (consumers, clinicians, managers, administrators, policy makers, and insurers).
- Adequate human and technical resources.

Each region planning to introduce a measurement system needs a cadre of committed leader-educators strongly supported by the top level of administration and provided with sufficient resources to implement the system. The educative process required to install and sustain the system is a continuous one, involving all stakeholders, educators included. Sustained commitment is required of leadership.

Glaser and Backer (1980) report that an innovative system will not endure unless it is administratively integrated into the host agency and congruent with the values of that agency. The installation of the system must be accompanied by open discussion with the staff of any problems as they arise, and continuous feedback to staff concerning the progress of the enterprise. The system is more likely to take root if clinicians find it valuable and if they derive benefits from it (e.g., information about patient progress that can be used to modify or refine initial treatment plans). The system should be sufficiently flexible to be adaptable to local needs and conditions (e.g., by being offered in elective, modular form). Finally, political circumstances applying at the outset can block installation, particularly if the staff are distracted or stressed (e.g., by a change of leadership or an industrial dispute). A favourable time should be chosen for introduction.

In our review of stakeholders' opinions (Chapter III), one professional group averred that they were "weary and wary" of change and that they would resist the introduction of an outcome measurement system. Block (1993) has discussed the means by which such opposition can be challenged. According to Block, cynics are not swayed by rational argument. The cynic's basic position is as follows: the organisation has been down the same path before; top management is not committed; and the innovator's tenure may be brief. The power of the cynic's position is twofold: there is truth in what he says; and he speaks for all of us. Other attempts at change have come and gone; top administrators do send mixed messages; and the innovator may be discarded, transferred or promoted sooner than he or she thinks. Cynics express everybody's doubts. The innovator, above all, is likely to have doubts about the validity of the innovation, about the risks involved, and about his or her capacity to bring it about. Because history is on their side and they express universal doubts, cynics cannot be argued out of their position or bartered with. Cynics lack faith in the sincerity of others. They ask for an assurance of certainty; but change is risky and certainty cannot be promised.

Block describes two other groups who are often allied to the cynics: "victims" and "bystanders". Victims want power without responsibility. They do not want a change in the authority structure, just a change in who is in charge. Bystanders withhold commitment. Before they will agree to collaborate, they demand incontestable proof that the innovation will work. However, other people's experience is no definite answer to the risk of changing one's own organisation. If innovation were a sure thing, no commitment would be required.

How can cynicism be neutralized and commitment fostered? Block says that faith, responsibility and commitment are matters of personal choice. The innovator must elicit the cynic's doubts, affirm the cynic's concern about the risks of change, acknowledge past administrative failures, and invite the cynic to join the innovative venture. Even if this approach fails to persuade, it will neutralize the opposition. The victim's feelings of helplessness are acknowledged, and bystanders' desire for more data supported. Critical in the process is the awareness that, despite their doubts, people cannot avoid choices. No choice is a choice. The strongest response to the cynic's cold water is the innovator's commitment to the idea that change is a matter of personal responsibility. The issue then becomes a choice that all the staff are invited to make. They will need time to make such a decision. At first, perhaps 25% will choose to get involved. That is enough. They will pull the others along, and the cynics will be left behind.

When key administrators are dubious or obstructive, the innovator promises operational outcomes in return for latitude of action. The case for innovation is presented, the need for exemption from normal controls acknowledged, and specific results promised. All that is needed is administrative tolerance (although enthusiasm would be preferable).

How should the innovation be introduced? The educational process should involve all stakeholders. The case for innovation should be presented by formal presentation backed up by succinct written material. The initial presentation could, for example, deal with the following matters:

- The need for accountability
- The rationale and purpose of outcome measurement
- The concept of a system
- Process and outcome
- The domains of measurement
- The measurement system recommended
- The particular measure, measures, or modules recommended
- Risks and opportunities
- Piloting the system

From the discussions that follow such presentations, enthusiasts, bystanders, and obstructionists will emerge, to be addressed in the manner already described.

Generally speaking, innovators should start simply, in a site where the staff (or most of them) express an interest in collaborating. Managers and staff can assist in the installation of the local system, and in discussing the problems that emerge as the system is piloted. Installation, testing and monitoring are a part of a planned collaborative process, the instrument being modified or refined in accordance with feedback from those who use it.

Further presentations are needed for all stakeholders on the following matters:

- Using goal-directed measurement to track patient progress and make treatment decisions

- Using concurrent and baseline and follow-up information to make administrative decisions
- The computerisation of measurement

Regular presentations are required concerning the progress of the enterprise, giving all stakeholders (cynics, victims and bystanders included), the opportunity to discuss and contribute to progress. Gradually, those who were originally neutral or obstructive are brought in or isolated. Mental health services will be empowered to embrace a spirit of enquiry and self-examination.

How Should these Recommendations be Implemented?

Given the above findings and issues we recommend that implementation of these measurement systems be incremental and evolutionary. The three modules (BFM, BM, and CM) should be developed simultaneously. A significant amount of psychometric work needs to be accomplished before any of the subsystems could be disseminated. According to stakeholder opinions, unless the Commonwealth can show that the measurement system is valid and useful, it is unlikely to be adopted. In terms of order of importance and ease of development we would suggest the following priority for developmental work.

1. Baseline-Followup Module
2. Concurrent Module
3. Background Module

Depending on the availability of funds, we estimate that initial development and field-testing would take three years. Starting in the third year we recommend that education and training be offered in how to use the measurement systems. We believe that the introduction of valid and useful measurement systems could empower child and youth mental health services to become learning organisations.

CHAPTER VI

REFERENCES

General References

Achenbach, T. M. (1991a). *Manual for the Child Behavior Checklist/4-18 and 1991 profile*. Burlington, VT: Department of Psychiatry, University of Vermont.

Achenbach, T. M. (1991b). *Manual for the Youth Self-Report and 1991 profile*. Burlington, VT: Department of Psychiatry, University of Vermont.

Achenbach, T. M., Hensley, V. R., Phares, V., & Grayson, D. (1990). Problems and competencies reported by parents of Australian and American children. *Journal of Child Psychology, Psychiatry and Allied Disciplines*, 31, 265-86.

Achenbach, T. M., McConaughy, S. H., & Howell, C.T. (1987). Child/adolescent behavioural and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, 101, 213-232.

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.

Anderson, J.G., & Jay, S.J. (1987). Computers and clinical medicine: implications for the practice of medicine. Pages 336-352 in J.G. Anderson & S.J. Jay (eds.), *Use and Impact of Computers in Clinical Medicine*. New York: Springer-Verlag.

Andrews, G., Peters, L., & Teesson, M. (1994). *The Measurement of Consumer Outcome in Mental Health: A report to the National Mental Health Information Strategy Committee*. Sydney: Clinical Research Unit for Anxiety Disorders.

Bagarozzi, D. A. (1986). Some issues to consider in the assessment of marital/family functioning. *American Journal of Family Therapy*, 14, 84-86.

Barkham, M., Rees, A., Stiles, W. B., Shapiro, D. A., Hardy, G. E. & Reynolds, S. (1996). Dose-effective relations in time-limited psychotherapy for depression. *Journal of Consulting and Clinical Psychology*, 64, 927-935.

Bickman, L. (Ed.). (1996). Special Issue: Methodological issues in evaluating mental health services. *Evaluation and Program Planning*, 19(2).

Bickman, L. (1997). Resolving issues raised by the Fort Bragg evaluation: New directions for mental health services research. *American Psychologist*, 52, 562-565.

Bickman, L., Lambert, E. W., & Schilling, S. (1998). Time patterns of mental health outcomes. *Mental Health Services Research*.

Bickman, L., Lambert, E. W., & Vides de Andrade, A. R. (in press). *The Long Term Outcomes of the Ft. Bragg Evaluation*.

Bickman, L., Rosof, J., Salzer, M., Summerfelt, W. T., Noser, K., Wilson, S. J., & Karver, M. S. (in press). What information do clinicians value for monitoring adolescent client progress. *Professional Psychology: Research and Practice*.

Biederman, J., Milberger, S., Faraone, S. V., Kiely, K., et al. (1995). Family-environment risk factors for attention-deficit hyperactivity disorder: A test of Rutter's indicators of adversity. *Archives of General Psychiatry*, 52, 464-470.

Bird, H. R., Canino, G. J., Rubio, S. M. & Ribera, J. C. (1987). Further measures of the psychometric properties of the Children's Global Assessment Scale. *Archives of General Psychiatry*, 44, 821-824.

Block, P. (1993). *Stewardship: Choosing service over self-interest*. San Francisco: Berrett-Koehler.

Bond, L., Nolan, T., Adler, R., & Robertson, C. (1994). The Child Behaviour Checklist in a Melbourne urban sample. *Australian Psychologist*, 29, 103-109.

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research, and Practice*, 16, 252-260.

Brann, P., & Wong, L. (1997). Developing a measure of how much we help children and adolescents. *HoNOSCA: The Health of the Nation Outcome Scales for Children and Adolescents*. Unpublished manuscript. Melbourne: Maroondah Hospital.

Brannan, A. M., Sonnichsen, S. E., & Heflinger, C. A. (1996). Measuring satisfaction with children's mental health services: Validity and reliability of the satisfaction scales. *Evaluation and Program Planning*, 19, 131-141.

Briffa, N. (1998). Patient data, confidentiality, and electronics [editorial]. *British Medical Journal*, 316:718-719.

Buri, J. R., Murphy, P., Richtsmeter, L. M., & Komar, K. K. (1992). Stability of parental nurturance as a salient predictor of self-esteem. *Psychological Reports*, 71, 535-543.

Byrne, B. M. (1983). Investigating measures of self-concept. *Measurement and Evaluation in Guidance*, 16, 115-125.

Carbonari, J. P., DiClemente, C. C., & Zweben, A. (1994, November 13). *A readiness to change measure*. Paper presented at the AABT National Meeting, San Diego, CA.

Carson, D. K., Swanson, D. M., Cooney, M. H., & Gillum, B. J. (1992). Stress and coping as predictors of young children's development and psychosocial adjustment. *Child Study Journal*, 22, 273-301.

Chappell, R., & Branch, L. G. (1993). The Goldilocks dilemma in survey design and its solution. *Journal of Clinical Epidemiology*, 46, 309-312.

Chamberlain, P., Ray, J., & Moore, K. J. (1996). Characteristics of residential care for adolescent offenders: A comparison of assumptions and practices in two models. *Journal of Child and Family Studies*, 5, 285-297.

Coddington, R. (1972). The significance of life events as etiologic factors in the diseases of children: A study of a normal population. *Journal of Psychosomatic Research*, 16, 205-403.

Commonwealth of Australia (1998). *The Privacy Act No.119*. Canberra, A.C.T.: Australian Government Publishing Service.

Compas, B. E., Davis, G. E., Forsythe, C. J., & Wagner, B. M. (1987). Assessment of major and daily stressful events during adolescence: The Adolescent Perceived Events Scale. *Journal of Consulting and Clinical Psychology*, 55, 534-541.

Cook, S. W., & Heppner, P. P. (1997). A psychometric study of three coping measures. *Educational and Psychological Measures*, 57, 906-923.

Cooper , H. & Hedges, L.V. (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.

Cournoyer, D. E., & Rohner, R. P. (1996). Reliability or retrospective reports of perceived maternal acceptance-rejection in childhood. *Psychological-Reports*, 78, 147-150.

Cross, T., & McDonald, E. (1995). *Evaluating the outcome of children's mental health services: A guide for the use of available child and family outcome measures*. Unpublished manuscript, Judge Baker Children's Center, Boston, MA.

Dawes, R. M. (1994). House of cards: *Psychology and psychotherapy built on myth*. New York: The Free Press.

Dew, T., & Huebner, E. S. (1994). Adolescents' perceived quality of life: An exploratory investigation. *Journal of School Psychology*, 32, 185-199.

Dornelas, E. A., Correll, R. E., Lothstein, L., & Wilber, C. (1996). Designing and implementing outcome evaluations: Some guidelines for practitioners. *Psychotherapy*, 33, 237-245.

Douglass, K. S., Macfarlane, E., & Webster, C. D. (1996). Predicting dangerousness in the contemporary Canadian mental health and criminal justice systems. *Canada's Mental Health*, 43, 4-11.

DuBois, D. L., Felner, R. D., Brand, S., Adan, A. M., & Evans, E. G. (1992). A prospective study of life stress, social support, and adaptation in early adolescence. *Child Development*, 63, 542-557.

Eisen, S. V., & Dickey, B. (1996). Mental health outcome assessment: The new agenda. *Psychotherapy*, 33, 181-189.

Eltz, M., Shirk, S., & Sarlin, N. (1995). Alliance formation and treatment outcome among maltreated adolescents. *Child Abuse and Neglect*, 19, 419-431.

Eyberg, S. M., & Robinson, E. A. (1983). Dyadic Parent-Child Interaction Coding System (DPICS): A manual. *Psychological Documents*, 13, (2) MS.2582:24

Flanagan, R. (1995). A review of the Behavior Assessment System for Children (BASC): Assessment consistent with requirements of the Individuals with Disabilities Education Act (IDEA). *Journal of School Psychology*, 33, 177-186.

Fobair, P. A., & Zabora, J. R. (1995). Family functioning as a resource variable in psychological cancer research: Issues and measures. *Journal of Psychosocial Oncology*, 13, 97-114.

Folkman, S., & Lazarus, R. S. (1985). If it changes it must be a process: Study of emotion and coping during three stages of college examination. *Journal of Personality and Social Psychology*, 48, 150-170.

Forman, B. D., & Hagan, B. J. (1983). A comparative review of total family functioning measures. *The American Journal of Family Therapy*, 11, 25-40.

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

Gammon, G., John, K., Prusoff, B. A., & Weissman, M. M. (1982). *Assessment of social adjustment in children and adolescents: Preliminary testing of a semistructured interview*. Unpublished manuscript, Yale University, New Haven, CT.

Garwick, G., & Lampman, S. (1972). Typical problems bringing patients to a community mental health center. *Community Mental Health Journal*, 8, 271-280.

Gilberts, R. (1983). The evaluation of self-esteem. *Family and Community Health*, 6, 9-49.

Glaser, E. M., & Backer, T. E. (1980). Durability of innovations : How goal attainment scaling programs fare over time. *Community Mental Health Journal*, 16, 130-143.

Goldman, B. A. & Mitchell, D. F. (1995). *Directory of unpublished experimental mental measures*, Vol. 6. Washington, DC: Amercian Psychological Association.

Goldman, B. A., Mitchell, D. F., Egelson, P.E. (1997). *Directory of unpublished experimental mental measures*, Vol. 7. Washington, DC: American Psychological Association.

Goodman, R. (1994). A modified version of the Rutter Parent Questionnaire including extra items on children's strengths: A research note. *Journal of Child Psychology and Psychiatry*, 35, 1483-1494.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Clinical Psychology and Psychiatry*, 38, 581-586.

Goodman, W. K., Price, L. H., Rasmussen, S. A., Riddle, M. A., & Rapoport, J. L. (1991). *Children's Yale-Brown Obsessive Compulsive Scale*. New Haven, CT : Child Study Center, Yale University School of Medicine.

Gorin, S. (1993). The prediction of child psychotherapy outcome: Factors specific to treatment. *Psychotherapy*, 30, 152-158.

Green, R. S., & Newman, F. L. (1996). Criteria for selecting instruments to assess treatment outcomes. *Residential Treatment for Children and Youth*, 13, 29-48.

Halvorsen, J. (1991). Self-report family assessment instruments: An evaluative review. *Family Practice Research Journal*, 11, 21-55.

Hammill, D. D., Brown, L., & Bryant, B. R. (1992). Evaluating the technical characteristics of tests, *A consumer's guide to tests in print* (2nd ed.). Austin, TX: Pro-Ed.

Harrison, P. L. (1989). Adaptive behavior: research to practice. *Journal of School Psychology*, 27, 301-317.

Headey, B., Kelley, J., & Wearing, A. (1993). Dimensions of mental health: Life satisfaction, positive affect, anxiety, and depression. *Social Indicators Research*, 29, 63-82.

Heflinger, C. A. (1992). Client-level outcomes of mental health services for children and adolescents. *New Directions for Program Evaluation*, 54, 31-45.

Heflinger, C. A., & Bickman, L. (1996). Family Empowerment: A conceptual model for promoting parent-professional partnership. In C. A. Heflinger & C. T. Nixon (Eds.), *Families and the mental health system for children and adolescents*. Thousand Oaks, CA.: Sage.

Hensley, V. R. (1988). Australian normative study of the Achenbach Child Behaviour Checklist. *Australian Psychologist*, 23, 371-382.

Hernandez, M., & Hodges, S. (1996). *The ecology of outcomes: System accountability in children's mental health*. Paper presented at the Proceedings of the 9th Annual Research Conference: A System of Care for Children's Mental Health: Expanding the Research Base, Tampa, FL.

Hines, A. M. (1993). Linking qualitative and quantitative methods in cross cultural survey research: Techniques from cognitive science. *American Journal of Community Psychology*, 21, 729-746.

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.

Hodges, K., & Wong, M. M. (1997). Use of the Child and Adolescent Functional Assessment Scale to predict service utilization and cost. *Journal of Mental Health Administration*, 24, 278-90.

Horvath, A. O. (1994). Research on the alliance. In A. O. Horvath & L. S. Greenberg (Eds.), *The working alliance: Theory, research, and practice* (pp. 259-286). New York: John Wiley & Sons.

Horvath, A. O., & Luborsky, L. (1993). The role of the therapeutic alliance in psychotherapy. *Journal of Consulting and Clinical Psychology*, 61, 561-573.

Huebner, E. S. (1991). Further validation of the Students' Life Satisfaction Scale: The independence of satisfaction and affect ratings. *Journal of Psychoeducational Assessment*, 9, 363-368.

Huebner, E. S. (1994). Conjoint analysis of the Students' Life Satisfaction Scale and the Piers-Harris Self-Concept Scale. *Psychology in the Schools*, 31, 273-277.

Huebner, E. S., & Dew, T. (1993). Is life satisfaction multidimensional? The factor structure of the Perceived Life Satisfaction Scale. *Journal of Psychoeducational Assessment*, 11, 345-350.

Hughes, D., & Dumont, K. (1993). Using focus groups to facilitate culturally anchored research. *American Journal of Community Psychology*, 21, 775-806.

Ivan, T. M., & Glazer, J. P. (1994). Quality of life in pediatric psychiatry: A new outcome measure. *Child and Adolescent Psychiatric Clinics of North America*, 3, 599-611.

Jellinek, M., Little, M., Murphy, J. M., & Pagano, M. (1995). The Pediatric Symptom Checklist. *Archives of Pediatric and Adolescent Medicine*, 149, 740-746.

Kazdin, A. E. (1980). Acceptability of alternative treatments for child deviant behavior. *Journal of Applied Behavior Analysis*, 13, 259-273.

Kazdin, A. E. (1993a). Psychotherapy for children and adolescents. *American Psychologist*, 48, 644-657.

Kazdin, A. E. (1993b). Changes in behavioral problems and prosocial functioning in child treatment. *Journal of Child and Family Studies*, 2, 5-22.

Kazdin, A., Holland, L., & Crowley, M. (1997). Family experience of barriers to treatment and premature termination from child therapy. *Journal of Consulting and Clinical Psychology*, 65, 453-463.

Kelley, M. L., Heffer, R. W., Gresham, F. M. & Elliott, S. N. (1989). Development of a modified Treatment Evaluation Inventory. *Journal of Psychopathology and Behavioral Assessment*, 11, 235-247.

Kiresuk, T. J., & Lund, S. H. (1982). Goal attainment scaling: A medical-correctional application. *Medicine and Law*, 1, 227-251.

Kiresuk, T. J. & Sherman, R. E. (1968). Goal attainment scaling : A general method for evaluating community mental health programs. *Community Mental Health Journal*, 4, 443-453.

Kiresuk, T. J., Smith, A., & Cardillo, J. E. (1994). *Goal attainment scaling: Applications, theory, and measurement*. Hillsdale, N.J.: Lawrence Erlbaum.

Klee, M. R., Abeles, N., & Muller, R. T. (1990). Therapeutic alliance: Early indicators, course, and outcome. *Psychotherapy*, 27, 166-174.

Kleinmuntz, B. (1990). Why we still use our heads instead of formulas: toward an integrative approach. *Psychological Bulletin*, 107, 296-310.

Knights, G. P., Virdin, L. M., & Roosa, M. (1994). Socialization and family correlates of mental health outcomes among Hispanic and Anglo American children: Consideration of cross-ethnic scalar equivalence. *Child Development*, 65, 212-224.

Knitzer, J. (1982). *Unclaimed children: The failure of public responsibility to children and adolescents in need of mental health services*. Washington, DC: Children's Defence Fund.

Knodel, J. (1993). In D.L. Morgan (Ed). *Successful focus groups: Advancing the state of the art*. California: SAGE.

LeBuffe, P. A., & Pfeiffer, S. I. (1996). Measuring outcomes in residential treatment with the Devereux Scales of Mental Disorders. *Residential Treatment for Children and Youth*, 13, 83-91.

Lehnhoff, J. (1993). Assessment and utilization of patient strengths in acute care treatment planning. *The Psychiatric Hospital*, 22, 11-15.

Long, K. A., & Hamlin, C. M. (1988). Use of the Piers-Harris Self-Concept Scale with Indian children: Cultural considerations. *Nursing Research*, 37, 42-46.

Luborsky, L. (1994). Therapeutic alliances as predictors of psychotherapy outcomes: Factors explaining the predictive success. In A. O. Horvath, & L. S. Greenberg (Eds.), *The working alliance: Theory, research, and practice* (pp. 38-50). New York: John Wiley & Sons.

Marsh, H. W., Carins, L., Relich, J., Barnes, J., & Debus, R.L. (1984). The relationship between dimensions of self-attribution and dimensions of self-concept. *Journal of Educational Psychology*, 76, 3-32.

Marsh, H. W., Relich, J. D., & Smith, I. D. (1983). Preadolescent self-concept: Their relationship to age, sex, and academic measures. *American Educational Research Journal*, 22, 422-444.

Marsh, H. W., Smith, I. D., & Barnes, J. (1983). Multi-trait multi-method analysis of the Self-description Questionnaire: Student-teacher agreement on multidimensional ratings of student self-concept. *American Educational Research Journal*, 20, 333-357.

McConnaughy, E. A., DiClemente, C. C., Prochaska, J. O., Velicer, W. F. (1989). Stages of changes in psychotherapy: A follow-up report. *Psychotherapy*, 26, 494-503.

McConnaughy, E. A., Prochaska, J. O., & Velicer, W. F. (1983). Stages of change in psychotherapy: Measurement and sample profiles. *Psychotherapy: Theory, Research, and Practice*, 20, 368-375.

Migneault, J. P., Pallonen,U. E., & Velicer, W. F.(1997). Decisional balance and stage of change for adolescent drinking, *Addictive Behaviours*, 22, 339-351

Miller, D. L., & Kelley, M. L. (1992). Treatment acceptability: The effects of parent gender, marital adjustment, and child behavior. *Child and Family Behavior Therapy*, 14, 11-23.

Milner, J., Gold, R., & Wimberly, R. (1986). Prediction and explanation of child abuse: Cross-validation of the Child Abuse Potential Inventory. *Journal of Consulting and Clinical Psychology*, 54, 865-866.

Minuchin, S. (1974). *Families and family therapy*. Cambridge, Mass: Harvard University Press.

Monahan, J. (1984). The prediction of violent behaviour: Toward a second generation of theory and policy. *American Journal of Psychiatry*, 141, 10-15.

Monahan, J. & Steadman, H. J. (1994). *Violence and mental disorder: developments in risk assessment*. Chicago: University of Chicago Press.

Moos, R., & Moos, B. (1981). *Family Environment Scale Manual*. Palo Alto, CA: Consulting Psychologists Press.

Murphy, J. M., & Jellinek, M. S. (1985). *Development of a brief psychosocial screening instrument for pediatric practice*. Final report. NIMH Contract No. 84M0213612. Rockville, MD: National Institute of Mental Health.

Narikawa, O., & Frith, S. (1972). *Measures of self-concept*. Los Angeles: Instructional Objectives Exchange.

National Health and Medical Research Council. *Aspects of Privacy in Medical Research*. Canberra, A.C.T.: Australian Government Publishing Service, Catalogue No. 905 5047 X.

National Mental Health Policy (1992). Canberra: Australian Government Publishing Service.

National Standards for Mental Health Services (1997). Canberra: Australian Government Publishing Service.

Najavits, L., M., & Strupp, H., H. (1994). Differences in the effectiveness of psychodynamic therapists: A process-outcome study. *Psychotherapy*, 31, 114-123.

Nelson-Gardell, D. (1997). Child report of treatment issue resolution: Pilot of a rapid assessment instrument. *Child Abuse and Neglect*, 21, 309-318.

Neuman, W. L. (1994). *Social Research Methods: Qualitative and Quantitative Approaches*. (2nd ed.). Boston: Allyn and Bacon.

Newman, F. L., Hunter, R. H., & Irving, D. (1987). Simple measures of progress and outcome in the evaluation of mental health services. *Evaluation and Program Planning*, 10, 209-218.

Nguyen, T. D., Attkisson, C. C., & Stegner B. L. (1983). Assessment of patient satisfaction: Development and refinement of a service evaluation questionnaire. *Evaluation and Program Planning*, 6, 229-314.

Nunnally, J. C. (1978). *Psychometric Theory*. (2nd ed.). New York: McGraw- Hill.

Nurcombe, B. (1987a). Diagnostic reasoning and treatment planning: I *Australian and New Zealand Journal of Psychiatry*, 21, 477-482.

Nurcombe, B. (1987b). Diagnostic reasoning and treatment planning: II. *Australian and New Zealand Journal of Psychiatry*, 21, 483-489.

Nurcombe, B. (1987c). Diagnostic reasoning and treatment planning: III. *Australian and New Zealand Journal of Psychiatry*, 21, 490-499.

Nurcombe, B. (1989). Goal-directed treatment planning and the principles of brief hospitalization. *Journal of the American Academy of Child and Adolescent Psychiatry*, 28, 26-30.

Nurcombe, B., & Gallagher, R. M. (1986) *The Clinical Process in Psychiatry*. New York: Cambridge.

Ogles, B. M., Davis, D. C., & Lunnen, B. S. (1998). *The Ohio Scales: A manual*. Unpublished manuscript. Athens, OH: Ohio University.

Orwin, R. G. (1994). Evaluating coding decisions. In H. Cooper & L. V. Hedges (Ed.), *The handbook of research synthesis* (pp. 139-162). New York: Russell Sage Foundation.

Packer, T., Race, K.E. H., & Hotch, D. F. (1994). Focus groups: a tool for consumer-based program evaluation in rehabilitation agency settings. *The Journal of Rehabilitation*, 60, 30-37.

Parfenoff, H., & Jose, P. E. (1989). *Measuring daily hassles in children*. Paper presented at the Biennial meeting of the Society for Research in Child Development, Kansas City, MO.

Parish, T. S. & Rankin, C. I. (1982). The Nonsexist Personal Attribute Inventory for Children: A report on its validity and reliability as a self-concept scale. *Educational and Psychological Measurement*, 42, 339-343.

Parish, T., & Taylor, J. (1978a). A further report on the validity and reliability of the Personal Attribute Inventory for Children as self-concept scale. *Educational and Psychological Measurement*, 38, 1225-1228.

Parish, T., & Taylor, J. (1978b). The Personal Attribute Inventory for Children: A report on its validity and reliability as a self-concept scale. *Educational and Psychological Measurement*, 38, 565-569.

Perosa, L. M., Hansen, J., & Perosa, S. L. (1981). The development of the Structural Family Interaction Scale. *Family Therapy*, 8, 77-90.

Perosa, S. L., & Perosa, L. M. (1990). Convergent and discriminant validity for family self-report measures. *Educational and Psychological Measurement*, 50, 855-868

Pfeiffer, S. I. (Ed.). (1996). *Outcome assessment in residential treatment*. New York: The Haworth Press.

Pfeiffer, S. I., & Shott, S. (1996) Implementing an outcome assessment project: logistical, practical, and ethical considerations. In S. I. Pfeiffer (Ed.), *Outcome assessment in residential treatment* (pp. 71-81). New York: Haworth Press.

Plapp, J., & Rey, J. (1994). Child and adolescent psychiatric services: Case audit and patient satisfaction. *Journal of Quality Clinical Practice*, 14, 51-56.

Pratt, S. I. & Moreland, K.L. (1996) Introduction to treatment outcome historical perspectives and current issues. In S.I. Pfeiffer (Ed.), *Outcome assessment in residential treatment*. New York: Haworth Press.

Prochaska, J. O., & DiClemente, C. C. (1986). The transtheoretical approach. In J. C. Norcross (Ed.), *Handbook of eclectic psychotherapy* (pp. 163-200). New York: Brunner/Mazel.

Prochaska, J. O., DiClemente, C. C., & Norcross, J. C. (1992). In search of how people change: Applications to addictive behaviours. *American Psychologist*, 47, 1102-1114.

Procidiano, M. E., & Heller, K. (1983). Measures of perceived social support from friends and from family: Three validation studies. *American Journal of Community Psychology*, 11, 1-24.

Rabinowitz, J. (1993). Diagnostic reasoning and reliability: a review of the literature and a model of decision-making. *The Journal of Mind and Behavior*, 14, 297-315.

Reilly, A. (1996, March). *Assessing functional strengths as well as impairments: Conceptual models and practical approaches*. Paper presented at the 9th annual meeting of the Florida Mental Health Institute's A System of Care for Children's Mental Health: Expanding the Research Base, Tampa, F.L.

Rey, J. M., Plapp, J. M., & Simpson P.S. (in press). Parental satisfaction and outcome: A four-year study in child and adolescent mental health service. *Australian and New Zealand Journal of Psychiatry*.

Reynolds, C. R., & Kamphaus, R. W. (1992). *Manual: Behavior Assessment System for Children*. Circle Pines, MN: American Guidance Service.

Robbins, J. M., & Taylor, L. (1997). *Development of the Adolescent Treatment Outcomes Module (ATOM)*. Paper presented at the 9th Annual Research Conference Proceedings, A System of Care for Children's Mental Health: Expanding the Research Base. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, Research and Training Center for Children's Mental Health.

Rock, D. L. & Bransford, J. D. (1992). An empirical evaluation of three components of the tetrahedron model of clinical judgment. *The Journal of Nervous and Mental Disease*, 180, 560-565.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Rossi, P. H., & Freeman, H. E. (1993). *Evaluation: A systematic approach* (5th ed.). (Vol. 12). Newbury Park, CA: Sage.

Rowlison, R. T., & Felner, R. D. (1988). Major life events, hassles, and adaptation in adolescence: Confounding in the conceptualization and measurement of life stress and adjustment revisited. *Journal of Personality and Social Psychology*, 55, 432-444.

Ruggeri, M. (1994). Patients' and relatives' satisfaction with psychiatric services: The state of the art of its measurement. *Social Psychiatry and Psychiatric Epidemiology*, 29, 212-227.

Schwab-Stone, M., Ayers, T., Kasprow, W., Voyce, C., Barone, C., Shriver, T., & Weissberg, R. (1995). No safe haven: A study of violence exposure in an urban community. *Journal of the American Academy of Child and Adolescent Psychiatry*, 34, 1343-1352.

Seaberg, J. R. (1988). Child Well-Being Scales: A critique. *Social-Work-Research-and-Abstracts*, 24, 9-15.

Sederer, L. I., Hermann, R., & Dickey, B. (1995). The imperative of outcome assessment in psychiatry. *American Journal of Medical Quality*, 10, 127-132.

Schaffer, D., Gould, M., 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.

Skinner, H. (1987). Self-report instruments for family assessment. In T. Jacob (Ed.), *Theories, methods, and findings in applied clinical psychology* (pp. 427-452). New York, NY: Plenum.

Smilkstein, G. (1978). The Family APGAR: A proposal for a family function test and its use by physicians. *Journal of Family Practice*, 6, 1231-1239.

Smith, D. C., Adelman, H. S., Nelson, P., Taylor, L., & Phares, V. (1987). Students' perceptions of control at school and problem behavior and attitudes. *Journal of School Psychology*, 25, 167-176.

Smith, G. R., Jr., Fischer, E. P., Nordquist, C. R., Mosley, C. L., & Ledbetter, N. S. (1997). Implementing outcomes management systems in mental health settings. *Psychiatric Services*, 48, 364-368.

Smith, G. R., Manderscheid, R. W., Flynn, L. M., & Steinwachs, D. M. (1997). Principles for assessment of patient outcomes in mental health care. *Psychiatric Services*, 48, 1033-1036.

Smith, G. R., Waxman, H., Snyder, D., & Raphael, M. (1996). Assessment tools in psychiatry. *American Journal of Medical Quality*, 11, S46-S49.

Snowden, L. R. (1996). Ethnic minority populations and mental health outcomes. In D. M. Steinwachs, L. M. Flynn, G. S. Norquist, & E. A Skinner (Eds.), Using client outcomes to improve mental health and substance abuse treatment. *New Direction for Mental Health Services*, 71, 79-87.

Sparrow, S. S. & Cicchetti, D. V. (1985). Diagnostic uses of the Vineland Adaptive Behavior Scales. *Journal of Pediatric Psychology*, 10, 215-224.

Spirito, A., Stark, L. J., & Williams, C. (1988). Development of a brief coping checklist for use with pediatric populations. *Journal of Pediatric Psychology*, 13, 555-574.

Standards Australia (1995). *Australian Standard AS 4400 – 1995: Personnel Privacy Protection in Healthcare Information Systems*. Homebush, N.S.Wales: Standards Australia.

Stedman, T., Yellowlees, P., Mellsop, G., Clarke, R., & Drake, S. (1997). *Measuring consumer outcomes in mental health: Field testing of selected measures of consumer outcomes in mental health*. ACT: Department of Health and Family Services.

Stewart, D. W., & Shamdasani, P. N. (1990). *Focus groups: theory and practice*. Newbury Park: Sage.

Stroul, B. A. (1996). *Children's mental health: Creating systems of care in a changing society*. Baltimore: Paul H. Brookes.

Strupp, H. H., & Hadley, S. W. (1977). A tripartite model of mental health and therapeutic outcomes. *American Psychologist*, 32, 187-196.

Summerfelt, W. T., Bickman, L. (1994). *Using level of functioning to measure outcome in children's mental health services research*. Nashville, TN: Vanderbilt University. VIPPS Occasional Paper

Taylor, L., Adelman, H., & Kaser-Boyd, N. (1986). The Origin Climate Questionnaire as a tool for studying psychotherapeutic process. *Journal of Child and Adolescent Psychotherapy*, 3, 10-16.

Trauer, T. (1998). Issues in the assessment of outcome in mental health. *Australian and New Zealand Journal of Psychiatry*, 32, 337-343.

Tobin, M., & Hickie, I. (1998). Outcomes focused service delivery: developing an academic-management partnership. *Australian and New Zealand Journal of Psychiatry*, 32, 327-336.

Tobin, D. L., Holroyd, K. A., Reynolds, R. V., & Wigal, J. K. (1989). The hierarchical factor structure of the Coping Strategies Inventory. *Cognitive Therapy and Research*, 13, 343-361.

Tryon, G. S., & Kane, A. S. (1993). Relationship of working alliance to mutual and unilateral termination. *Journal of Counseling Psychology*, 40, 33-36.

Vaidya, N. (1996). Behavior Dimensions Rating Scale: A review. Measurement and Evaluation in Counseling and Development, 29, 56-60.

Waller A.A. (1991). Legal aspects of computer-based patient records and records systems. Pages 156-179 in R.S. Dick & E.B. Steen (eds.), *The Computer-Based Patient Record*. Washington, National Academy Press.

Wasserstein, S. B., & La Greca, A. M. (1996). Can peer support buffer against behavioral consequences of parental discord? *Journal of Clinical Child Psychology*, 25, 177-182.

Wedding, D., & Faust, D. (1989). Clinical judgment and decision making in Neuropsychology. *Archives of Clinical Neuropsychology*, 4, 233-265.

Weed, L. L. (1968). Medical records that guide and teach. *New England Journal of Medicine*, 278, 593-600.

Wiekhorst, C. L. (1974). Relationships between sources of school-related pressure and academic achievement. *Dissertation Abstracts International*, 34, 7601-7602.

Wolfe, V. V., Gentile, C., & Wolfe, D. A. (1989). The impact of sexual abuse on children: A PTSD formulation. *Behavior Therapy*, 20, 215-228.

Zimmerman, D. P. (1996). A comparison of commonly used treatment measures. In S.I. Pfeiffer (Ed.), *Outcome assessment in residential treatment* (pp. 49 - 69). New York: The Haworth Press.

Zubrick, S. R., Silburn, S. R., Garton, A., Burton, P., Dalby, R., Carlton, J., Shepherd, C., & Lawrence, D (1995). *Western Australian Child Health Survey: Developing Health and Wellbeing in the Nineties*. Perth, WA: Australian Bureau of Statistics and the Institute for Child Health Research.

References to Measurement Instruments

Coping Skills

Adolescent-Coping Orientation for Problem Experiences Scale

Patterson, J. M., & McCubbin, H. I. (1987). Adolescent coping styles and behaviors: conceptualization and measurement. *Journal of Adolescence*, 10, 163-186.

Adolescent Coping Scale

Frydenberg, E., & Lewis, R. (1993). Boys play sport and girls turn to others: Age, gender and ethnicity as determinants of coping. *Journal of Adolescence*, 16, 253-266.

Analysis of Coping Style

Boyd, H. F., & Johnson, G. O. (1981). *Analysis of coping style*. New York: Psychological Corporation.

Coping Inventory for Adolescents

Fanshawe, P. G., & Burnett, P. C. (1991). Assessing school-related stressors and coping mechanisms in adolescents. *British Journal of Educational Psychology*, 61, 92-98.

Coping Responses Inventory

Moos, R. H. (1997). Coping Responses Inventory: A measure of approach and avoidance coping skills. In C. P. Zalaquett & R. J. Wood (Eds.), *Evaluating stress: A book of resources* (Vol. 17, pp. 51-65). Lanham, MD: Scarecrow Press.

Coping Strategies Inventory

Tobin, D. L., Holroyd, K. A., & Reynolds, R. V. C. (1984). *Manual for the coping strategies inventory*. Unpublished manuscript, Ohio University, Athens.

Kidcope

Spirito, A., Stark, L. J., & Williams, C. (1988). Development of a brief coping checklist for use with pediatric populations. *Journal of Pediatric Psychology*, 13, 555-574.

Problem Inventory for Adolescent Girls

Gaffney, L. R. (1984). A multiple-choice test to measure social skills in delinquent and nondelinquent adolescent girls. *Journal of Consulting and Clinical Psychology*, 52, 911-912.

Social Coping Questionnaire for Gifted Students

Swiatek, M. A. (1995). An empirical investigation of social coping strategies used by gifted adolescents. *Gifted Child Quarterly*, 39, 154-161.

Stress Response Scale

Chandler, L. A. (1984). Behavioral responses of children to stress. In J. H. Humphrey (Ed.), *Stress in childhood* (pp. 47-61). New York: AMS Press.

Ways of Coping Scale

Folkman, S., & Lazarus, R. S. (1985). If it changes it must be a process: Study of emotion and coping during three stages of college examination. *Journal of Personality and Social Psychology*, 48, 150-170.

Family Functioning

Adolescent Family Life Satisfaction Index

Henry, C. S., Ostrander, D. L., & Lovelace, S. G. (1992). Reliability and validity of the Adolescent Family Life Satisfaction Index. *Psychological Reports*, 70, 1223-1229.

Caregiver Strain Questionnaire (formerly Burden of Care Questionnaire)

Brannan, A. M., Heflinger, C. A., & Bickman, L. (1997). The Caregiver Strain Questionaire: Measuring the impact on the family of living with a child with serious emotional disturbance. *Journal of Emotional & Behavioral Disorders*, 5, 212-222.

Child and Adolescent Burden Assessment

Messer, S. C., Angold, A., Costello, E. J., & Burns, B. (1996). The Child and Adolescent Burden Assessment (CABA): Measuring the family impact of emotional and behavioral problems. *International Journal of Methods in Psychiatric Research*, 6, 261-284.

Children’s Report of Parental Behavior Inventory

Schaefer, E. S. (1965). Children's reports of parental behavior: *An inventory. Child Development*, 36, 413-424.

Cleminshaw-Guidubaldi Parent Satisfaction Scale

Guidubaldi, J., & Cleminshaw, H. K. (1985). The development of the Cleminshaw-Guidubaldi Parent Satisfaction Scale. *Journal of Clinical Child Psychology*, 14, 293-298.

Clinical Rating Scale for the Circumplex Model of Marital and Family Systems

Olson, D. H., & Killorin, E. (1985). *Chemically dependent families and the circumplex model*. Unpublished manuscript. St. Paul: Department of Family Social Science, University of Minnesota.

Dyadic Parent-Child Interaction Coding System

Eyberg, S.M. (1974). *Manual for coding dyadic parent-child interactions*. Unpublished manuscript.

Family Adaptability and Cohesion Scales I

Olson, D. H., Bell, R., & Portner, J. (1978). *FACES: Family Adaptability and Cohesion Evaluation Scales* (Manual). St. Paul: Department of Family Social Science, University of Minnesota.

Family Adaptability and Cohesion Scales II

Olson, D. H., Bell, R & Portner, J., (1982). FACES II: Family Adaptability and Cohesion Evaluation Scales. In D. H. Olson, H. McCubbin, H. Barnes, A. Larsen, M. Musen, & M. Wilson (Eds.), *Family inventories* (pp. 5-24). St. Paul: Department of Family Social Science, University of Minnesota.

Family Adaptability and Cohesion Scales III

Olson, D. H., Portner, J., & Lavee, Y. (1985). *FACES III*. (Available from Family Social Science, 290 McNeal Hall, University of Minnesota, St. Paul, MN 55108).

Family APGAR

Smilkstein, G. (1978). The Family APGAR: A proposal for a family function test and its use by physicians. *Journal of Family Practice*, 6, 1231-1239.

Family Assessment Measure III

Skinner, H. A., Steinhauer, P. D., & Santa-Barbara, J. (1983). The Family Assessment Measure. *Canadian Journal of Community Mental Health*, 2, 91-105.

Family Environment Scale

Moos, R., & Moos, B. (1981). *Family Environment Scale Manual*. Palo Alto, CA: Consulting Psychologists Press.

Family Expressiveness Questionnaire

Greenberg, M., Mason, C., Lengua, L., & Conduct Problems Prevention Research Group. (1995). *Family Expressiveness Questionnaire: Technical report*. Unpublished manuscript. University of Washington, Seattle.

Family Functioning Index

Pless, I. B., & Satterwhite, B. (1973). A measure of family functioning and its application. *Social Science Medicine*, 7, 613.

Family Functioning Questionnaire

Linder-Pelz, S., Levy, S., Tamir, A., Spenser, T., & Epstein, L. M. (1984). A measure of family functioning for health care practice and research. *Israel Journal of Comparative Family Studies*, 15, 211-230.

Family Functioning Scales

Bloom, B. L. (1985). A factor analysis of self-report measures of family functioning. *Family Process*, 24, 225-231.

Family Functioning Style Scale

Dunst, C. J., Trivette, C. M., & Deal, A. G. (Eds.). (1988). *Enabling and empowering families: Principles and guidelines for practice*. Cambridge, MA: Brookline.

Family Intervention Scale

Taynor, J., Nelson, R., & Daugherty, W. K. (1990). The Family Intervention Scale: Assessing treatment outcome. *Families in Society*, 71, 202-210.

Family Invulnerability Test

McCubbin, H. I., Olson, D. H., Lavee, Y., & Patterson, J. M. (1985). *Family Invulnerability Test: Stresses, strengths and adaptation*. St. Paul: Department of Family Social Science, University of Minnesota.

Family Life Questionnaire

Lautrey, J. (1980). *Classe sociale, milieu familial, intelligence*. Paris: PUF

Family Relationship Inventory

Michaelson, R. B., & Bascom, H. (1978). *Family Relationship Inventory*. Los Angeles: Psychological Publications.

Family Relationships Inventory

Bachman, J. G. (1970). The impact of family background and intelligence on tenth grade boys. In *Youth in transition* (Vol. 1). Ann Arbor, MI: Institute for Social Research.

Family Satisfaction

Olson, D. H., & Wilson, M. (1985). Family satisfaction. In D. H. Olson, H. I. McCubbin, H. Barnes, A. S. Larsen, M. Musen, & M. Wilson. (Eds.), *Family inventories* (pp. 43-50). St. Paul, MN: University of Minnesota.

Family Stressor Scale

Kessler, R. C. (1985). *Family Stressor Scale* (used in a 1985 Detroit area survey). Unpublished questionnaire, Ann Arbor, MI: Institute for Social Research.

Global Family Environment Scale

Rey, J. M., Singh, M., Hung, S., Dossetor, D. R., Newman, L., Plapp, J. M., & Bird, K. D. (1997). A global scale to measure the quality of family environment. *Archives of General Psychiatry*, 54, 817-822.

Issues Checklist

Robins, A. L., & Weiss, J. G. (1980). Criterion-related validity of behavioral and self-report measures of problem-solving communication skills in distressed and non-distressed parent-adolescent dyads. *Journal of Counseling Psychology*, 35, 203-208.

McMaster Family Assessment Device

Epstein, N. B., Baldwin, L. M., & Bishop, D. S. (1983). The McMaster Family Assessment Device. *Journal of Marital and Family Therapy*, 9, 171-180.

Modified Impact on Family Scale

Stein, R. E. K., & Riessman, C. K. (1980). The development of an impact-on-family scale: Preliminary findings. *Medical Care*, 18, 465-472.

O’Leary-Porter Scale

Porter, B., & O'Leary, K. D. (1980). Marital discord and childhood behavior problems. *Journal of Abnormal Child Psychology*, 8, 287-295.

Parent Perception Inventory

Hazzard, A., Christensen, A., & Margolin, G. (1983). Children's perceptions of parental behaviors. *Journal of Abnormal Child Psychology*, 11, 49-60.

Parent-Adolescent Attachment Inventory

Armsden, G. C., & Greenberg, M. T. (1987). The Inventory of Parent and Peer Attachment: Relationships to well-being in adolescence. *Journal of Youth and Adolescence*, 16, 427-454.

Parent-Adolescent Communication Inventory

Barnes, H., & Olson, D. H. (1982). Parent-adolescent communication. In D. H. Olson, H. I. McCubbin, H. Barnes, A. Larson, M. Musen, & M. Wilson (Eds.), *Family inventories*. St. Paul, MN: Family Social Science, University of Minnesota.

Parent-Adolescent Relationship Scale

Bell, N.J. & Avery, A.W. (1987) Family structure and parent-adolescent relationships: Does family structure really make a difference? *Journal of Marriage and Family*, (47) 2, 503-508.

Parental Acceptance-Rejection Questionnaire

Rohner, R. P., Saavedra, J. M., & Granum, E. O. (1978). Development and validation of the Parental Acceptance-Rejection Questionnaire: Test manual. *JSAS Catalog of Selected Documents in Psychology* (manuscript 1635), 8, 7-8.

Parental Nurturance Scale

Buri, J. (May 1989). *An instrument of parental authority prototypes*. Paper presented at the Sixty-First Annual Meeting of the Midwestern Psychological Association, Chicago.

Parent-Child Relationship Inventory

Gerard, A. B. (1994). *Parent-child relationship inventory*. Los Angeles, CA : Western Psychological Services.

Parent-Child Scales

Hetherington, E. M., & Clingempeel, W. G. (1992). Coping with marital transitions. *Monographs of the Society in Child Development*, 57, 1-238.

Parenting Stress Index

Abidin, R. R. (1997). Parenting Stress Index: A measure of the parent-child system. In C.P. Zalaquett, & R.J. Wood, *Evaluating stress: A book of resources* (pp.277-291). Lanham, MD, Scarecrow Press.

Perceptual Indicators of Family Quality of Life Scale

Rettig, K. D. et al. (1989). *Family life quality: Theory, assessment, and replication*. Paper presented at the preconference workshop on Theory Construction and Research Methodology, National Council on Family Relations, New Orleans, LA.

Quality of Parent-Child Relationships scale (author unknown)

Self-Report Family Inventory

Beavers, W. R., Hampson, R. B., & Hulgus, Y. F. (1985). Commentary: The Beavers systems approach to family assessment. *Family Process*, 24, 398-405.

Structural Family Interaction Scale-Revised

Perosa, L. M., Hansen, J., & Perosa, S. L. (1981). The development of the Structural Family Interaction Scale. *Family Therapy*, 8, 77-90.

Youth-Parent Conflict Scale

Prinz, R. J. (1979). The conflict behavior questionnaire. In A. L. Rubin, & S. C. Foster (Eds.), *Negotiating parent-adolescent conflict* (pp. 299-308). New York: Guilford.

Family Resources

Family Needs Scale

Dunst, C. J., Trivette, C. M., & Deal, A. G. (Eds.). (1988). *Enabling and empowering families: Principles and guidelines for practice*. Cambridge, MA: Brookline.

Family Needs Survey

Bailey, D. B., & Simeonsson, R. J. (1985). *Family Needs Survey*. Chapel Hill, NC: Frank Porter Graham Child Development Center, University of North Carolina.

Family Resource Scale

Dunst, C. J., & Leet, H. E. (1987). Measuring the adequacy of the resources in households with young children. *Child: Care, Health, and Development*, 13, 111-125.

FISC Family Needs Survey

McGrew, K. S., Gilman, C. J., & Johnson, S. D. (1989). *Family Needs Survey results: Responses from parents of young children with disabilities*. Minneapolis, MN: University of Minnesota, Institute on Community Integration.

Home Index

Gough, H. G. (1949). A short social status inventory. *Journal of Educational Psychology*, 52-56.

Prioritizing Family Needs Scale

Finn, D. M., & Vadasy, P. F. (1988). *Prioritizing Family Needs Scale*. Seattle: University of Washington.

Support Functions Scale

Dunst, C. J., Trivette, C. M., & Deal, A. G. (Eds.). (1988). *Enabling and empowering families: Principles and guidelines for practice*. Cambridge, MA: Brookline.

Functioning Competence

Adaptive Behaviour Inventory for Children

Mercer, J., & Lewis, J. (1982). *Adaptive Behavior Inventory for Children*. San Antonio, TX: Psychological Corporation.

Assessment of Interpersonal Relations

Bracken, B. A., & Howell, K. K. (1991). Multidimensional self-concept validation: A three-instrument investigation. *Journal of Psychoeducational Assessment*, 9, 319-328.

Behavioral and Emotional Rating Scale

Epstein, M. H., & Sharma, J. M. (1998). *Behavioral and Emotional Rating Scale: A strength-based approach to assessment*. Austin, TX: Pro-Ed.

CBCL and YSR Social Competence Scale

Achenbach, T. M. (1991a). *Manual for the Child Behavior Checklist/4-18 and 1991 profile*. Burlington, VT: Department of Psychiatry, University of Vermont.

Achenbach, T. M. (1991b). *Manual for the Youth Self-Report and 1991 profile*. Burlington, VT: Department of Psychiatry, University of Vermont.

Child Behavior Rating Scale

Rochester Social Problem Solving Core Group. (1980). *The Child Behavior Rating Scale*. Unpublished manuscript. University of Rochester, Rochester, NY.

Health Resources Inventory

Gesten, E. L. (1976). A health resources inventory: The development of a measure of the personal and social competence of primary-grade children. *Journal of Consulting and Clinical Psychology*, 44, 775-786.

Instrumental and Social Competence Scale

Beiser, M., Lancee, W., Gotowiec, A., Sack, W., & Redshirt, R. (1993). Measuring self-perceived role competence among First Nations and non-native children. *Canadian Journal of Psychiatry*, 38, 412-419.

Normative Adaptive Behavior Checklist

Adams, G.L. (1984). *Normative Adaptive Behavior Checklist*. San Antonio, TX: Psychological Corporation.

Responsibility and Independence Scale for Adolescents

Salvia, J., Neisworth, J. & Schmidt, M. (1990). *Responsibility and Independence Scale for Adolescents*. Allen, TX:DLM.

Vineland Adaptive Behavior Scales

Sparrow, S. S., Balla, D., & Cichetti, D. (1984). *Vineland Adaptive Behavior Scales*. Circle Pines, MN: American Guidance Service.

Functioning Impairment

Child and Adolescent Functional Assessment Scale

Hodges, K., Kline, J., Stern, L. Cytryn, L., & McKnew, D. (1982). The development of a child assessment interview for research and clinical use. *Journal of Abnormal Child Psychology*, 10, 173-189.

Children's Global Assessment Scale

Schaffer, D., Gould, M., 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.

Columbia Impairment Scale

Bird, H. R., Schaffer, D., Fisher, P., Gould, M., Staghezza, B., Chen, J.Y. & Hoven, C. (1993). The Columbia Impairment Scales (CIS): Pilot findings on a measure of global impairment for children and adolescents. *International Journal of Methods of Psychiatric Research*, 3, 167-176.

Hampstead Child Adaptation Measure

Fonagy, P., & Target, M. (1996). Predictors of outcome in child psychoanalysis: A retrospective study of 763 cases at the Anna Freud Centre. *Journal of the American Psychoanalytic Association*, 44, 27-77.

Personal Problems Checklist-Adolescent (PPC-A) and Children's Problems Checklist (CPC)

Schinka, J. (1985). *Children's Problems Checklist*. Odessa, FL: Psychological Assessment Resources.

Schinka, J. A. (1985). *Personal Problems Checklist - Adolescent*. Odessa, FL: Psychological Assessment Resources.

Teacher Observation of Classroom Adaptation -Revised (TOCA-R)

Werthamer-Larsson L., Kellam, S.G., Wheeler , L. (1991). Effect of first-grade classroom environment on shy behavior, aggressive behavior and concentration problems. *American Journal of Community Psychology*, 19, 585-602.

General Symptoms

Achenbach scales: Child Behavior Checklist 4/18, Youth Self-Report, Teacher Report Form

Achenbach, T. M. (1991a). *Manual for the Child Behavior Checklist/4-18 and 1991 profile*. Burlington, VT: Department of Psychiatry, University of Vermont.

Achenbach, T. M. (1991b). *Manual for the Youth Self-Report and 1991 profile*. Burlington, VT: Department of Psychiatry, University of Vermont.

Achenbach, T. M. (1991c). *Manual for the Teacher's Report Form and 1991 Profile*. Burlington, VT: Department of Psychiatry, University of Vermont

ACQ Behavior Checklist

Achenbach, T. M., Conners, C. K., & Quay, H. C. (1983). *The ACQ Behavior Checklist*. Burlington, VT: Department of Psychiatry, University of Vermont.

Behavior Dimensions Rating Scale

Bullock, L. M., & Wilson, M. J. (1989). *Examiner's manual: Behavior Dimensions Rating Scale*. Allen, TX: DLM Teaching Resources.

Behavioral Problems Index

Peterson, J. L., & Zill, N. (1986). Marital disruption, parent-child relationships, and behavior problems in children. *Journal of Marriage and the Family*, 48, 295-307.

Behavior Rating Profile-2nd edition

Brown, L., & Hammill, D. D. (1990). *Behavior Rating Profile* (2nd ed.). Austin, TX: Pro-Ed.

Bristol Social Adjustment Guides

Stott, D. H. (1969). *The social adjustment of children*. London: University of London Press.

Burks' Behavior Rating Scales

Burks, H. F. (1977). *Burks Behavior Rating Scales: Manual*. Los Angeles: Western Psychological Services.

Devereux Scales of Mental Disorders

Naglieri, J. A., LeBuffe, P. A., & Pfeiffer, S. I. (1994). *The Devereux Scales of Mental Disorders*. San Antonio, TX: Psychological Corporation

Eyberg Child Behavior Inventory

Eyberg, S. M., & Ross, A. W. (1978). Assessment of child behavior problems: The validation of a new industry. *Journal of Clinical Child Psychology*, 7, 113-116.

Louisville Behavioral Checklist

Miller, L. C., & Roid, G. H. (1988). Factor-analytically derived scales for the Louisville Behavior Checklist. *Journal of Consulting and Clinical Psychology*, 56, 302-304.

Ontario Child Health Study scales

Boyle, M. H., Offord, D. R., Racine, Y., Felming, J. F., Szatmari, P., & Sanford, M. (1993). Evaluation of the revised Ontario Child Health Study Scales. *Journal of Child Psychology and Psychiatry*, 34, 189-213.

Parent Daily Report

Chamberlain, P. (1980). *Standardization of a parent report measure*. Unpublished doctoral dissertation, University of Oregon, Eugene, Oregon.

Pediatric Symptom Checklist

Murphy, J. M., & Jellinek, M. S. (1985). *Development of a brief psychosocial screening instrument for pediatric practice*. Final report. NIMH Contract No. 84M0213612. Rockville, MD: National Institute of Mental Health.

Revised Behavior Problem Checklist

Quay, H. C. (1983). A dimensional approach to Behavior Disorder: The revised Behavior Problem Checklist. *School Psychology Review*, 12, 244-249.

Rutter's Parent Questionnaire

Rutter, M., Tizard, J., & Whitmore, K. (1970). *Education, Health, and Behavior*. London: Longman.

Vermont System for Tracking Client Progress

Burchard, J., & Bruns, E. (1993). *User's guide to the Weekly Adjustment Indicator Checklist: Vermont System for Tracking Client Progress*. Burlington, VT: Department of Psychology, University of Vermont.

Walker Problem Behavior Identification Checklist

Walker, H. M. (1976). *Walker Problem Behavior Identification Checklist Manual*. Los Angeles: Western Psychological Services.

Goals

Adolescent Goal Attainment Scale

Maher, C.A., & Barbrack, C.R. (1984). Evaluating individual counseling of conduct problem adolescents: The goal attainment scaling method. *Journal of School Psychology*, 22 (3), 285-297

Family Goal Recording

Fleuridas, C., & Rosenthal, D.M., Leigh, G.K., Leigh, T.E. (1990). Family goal recording: An adaptation of goal attainment scaling for enhancing family therapy and assessment. *Journal of Marital and Family Therapy*, 16, 389-406.

Maltreatment

Adolescent Abuse Inventory

Sebes, J.M. (1983). *Determining risk for abuse in families with adolescents: The development of a criterion measure*. Unpublished doctoral dissertation, Pennsylvania State University.

Checklist for Child Abuse Evaluation

Petty, J. (1990). *Checklist for Child Abuse Evaluation*. Odessa, FL: Psychological Assessment Resources.

Child Abuse Potential Inventory

Milner, J., Gold, R., & Wimberly, R. (1986). Prediction and explanation of child abuse: Cross-validation of the Child Abuse Potential Inventory. *Journal of Consulting and Clinical Psychology*, 54, 865-866.

Childhood Trauma Questionnaire

Bernstein, D. P., Fink, L., Handelsman, L., Foote, J., et al. (1994). Initial reliability and validity of a new retrospective measure of child abuse and neglect. *American Journal of Psychiatry*, 151, 1132-1136.

Children’s Impact of Traumatic Events Scale – R

Wolfe, V. V., Gentile, C., Michienzi, T., & Sas, L., et al. (1991). The Children's Impact of Traumatic Events Scale: A measure of post-sexual-abuse PTSD symptoms. *Behavioral Assessment*, 13, 359-383.

Child Report of Treatment Issue Resolution

Nelson-Gardell, D. (1997). Child report of treatment issue resolution: Pilot of a rapid assessment instrument. *Child Abuse and Neglect*, 21, 309-318.

Child Sexual Behavior Inventory

Friedrich, W. (1990). *Psychotherapy of sexually abused children and their families*. New York: W.W. Norton & Company.

History of Victimization Form

Gentile, C. (1988). *Factors mediating the impact of child-sexual abuse: Learned helplessness and severity of abuse*. Unpublished master’s thesis. University of Western Ontario, London, Ontario, Canada.

Sexual Abuse Fear Evaluation Subscale

Wolfe, V. V., & Wolfe, D. A. (1986). *The Sexual Abuse Fear Evaluation (SAFE): A subscale for the fear survey schedule for children - revised*. Unpublished questionnaire. University of Western Ontario, London, Ontario, Canada.

Trauma Symptom Checklist

Briere, J., & Runtz, M. (1989). The Trauma Symptom Checklist (TSC-33): Early data on a new scale. *Journal of Interpersonal Violence*, 4, 151-163.

Multidimensional

Adolescent Treatment Outcomes Module

Robbins, J. M., Taylor, J. L., Burns, B. J., Rost, K. M., Kramer, T., & Smith, G. R. (1997). *Adolescent Treatment Outcomes Module (ATOM): User's manual*. Little Rock, AR: University of Arkansas for Medical Sciences.

Behavior Assessment System for Children

Reynolds, C. R., & Kamphaus, R. W. (1992). *Manual: Behavior Assessment System for Children*. Circle Pines, MN: American Guidance Service.

Child Behavior Rating Scale

Kirkpatrick, D. (1975). *The effect of sex of rater in adult rating of child behavior*. Unpublished doctoral dissertation, University of Victoria, Victoria, B. C., Canada.

The Health of the Nation Outcome Scales for Children and Adolescents – Version 6.0; (HoNOSCA)

Gowers, S., Whitton, A. & Harrington, R. (1998). *HONOSCA: Health of the Nation Outcome Scales. Child and Adolescent version*. Report on research and development. University of Manchester, September 1995 – March 1997.

Life Stressors and Social Resources Inventory-Youth

Moos, R. H., Fenn, C. B., Billings, A. G., & Moos, B. I. (1989). Assessing life stressors and social resources: Applications to alcoholic patients. *Journal of Substance Abuse*, 1, 135-152.

Matson Evaluation of Social Skills with Youngsters

Matson, J. L., Rotatori, A. R., & Helsel, W. J. (1983). Development of a rating scale to measure social skills in children: The Matson Evaluation of Social Skills with Youngsters. *Behavior Research Therapy*, 21, 335-340.

Mental Status Checklist

Dougherty, E. H., & Schinka, J. A. (1988). *Mental Status Checklist - Adolescent*. Odessa, FL: Psychological Assessment Resources.

Dougherty, E. H., & Schinka, J. A. (1989). *Mental Status Checklist - Children*. Odessa, FL: Psychological Assessment Resources.

Modified Rutter A2 Parent Questionnaire

Goodman, R. (1994). A modified version of the Rutter Parent Questionnaire including extra items on children's strengths: A research note. *Journal of Child Psychology and Psychiatry*, 35, 1483-1494.

Ohio Youth Problems, Functioning, and Satisfaction Scales

Ogles, B. M. Davis, D. C. & Lunnen, B. S. (1998). *The Ohio Scales: A manual*. Unpublished manuscript. Athens, OH: Ohio University.

Paddington Complexity Scale

Hunter, J., Higginson, I., & Garralda, M.E. (1996). Systematic literature review: outcome measures for child and adolescent mental health services. *Journal of Public Health Medicine*, (18), 197-206

Questionnaire developed for specific study

Slee, P., & Rigby, K. (1993). Australian school children's self appraisal of interpersonal relations: The bullying experience. *Child Psychiatry and Human Development*, 23, 273-282.

Scales of Independent Behavior

Bruininks, R. K., Woodcock, R. W., Weatherman, R. F., & Hill, B. K. (1984). *Scales of Independent Behavior* (SIB). Chicago IL: Riverside Publishing.

Service Utilization and Risk Factors Interview

Goodman, S. H., Hoven, C. W., Narrow, W. E., Cohen, P., Fielding, B., Alegria, M., Leaf, P. J., Kandel, D., Horwitz, S. M., Bravo, M., Moore, R., & Dulcan, M. K. (1998). Measurement of risk for mental disorders and competence in a psychiatric epidemiologic community survey: The NIMH Methods for the Epidemiology of Child and Adolescent Mental Disorders (MECA) Study. *Social Psychiatry*, 33 (4), 162-173

Social Adjustment Inventory for Children and Adolescents

Gammon, G., John, K., Prusoff, B. A., & Weissman, M. M. (1982). *Assessment of social adjustment in children and adolescents: Preliminary testing of a semistructured interview*. Unpublished manuscript, Yale University, New Haven, CT.

Social Skills Rating System

Gresham, F. M., & Elliot, S. (1990). *Manual for the Social Skills Rating System*. Circle Pines, MN: American Guidance Service.

Strengths and Difficulties Questionnaire

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 38, 581-586.

Timberlawn Child Functioning Scale

Dimperio, T. L., Blotcky, M. J., Gossett, J. T., & Doyle, A. H. (1986). The Timberlawn Child Functioning Scale: A preliminary report on reliability and validity. *The Psychiatric Hospital*, 17, 115-120.

Quality of Life

Child Health and Illness Questionnaire

Starfield, B., Bergner, M., Ensminger, M., et al. (1993). Adolescent health status measurement: development of the child health and illness profile. *Pediatrics*, 91, 430-450.

Child Well-Being Scales

Magura, S., & Moses, B. S. (1986). *Outcome measures for clinical welfare services*. Washington, D.C.: Child Welfare League of America.

Hopelessness Scale for Children

Kazdin, A. E., & Petti, T.A. (1982). Self-report and interview measures of childhood and adolescent depression. *Journal of Psychology and Psychiatry*, 23, 437-457.

Perceived Life Satisfaction Scale

Smith, D. C., Adelman, H. S., Nelson, P., Taylor, L., & Phares, V. (1987). Students' perceptions of control at school and problem behavior and attitudes. *Journal of School Psychology*, 25, 167-176.

Quality of Life Profile: Adolescent version

Raphael, D., Rukholm, E., Brown, I., Hill-Bailey, P., & Donato, E. (1996). The Quality of Life Profile-Adolescent Version: Background, description, and initial validation. *Journal of Adolescent Health*, 19, 366-375.

Students' Life Satisfaction Scale

Huebner, E. S. (1991). Further validation of the Students' Life Satisfaction Scale: The independence of satisfaction and affect ratings. *Journal of Psychoeducational Assessment*, 9, 363-368.

Quality of School Life Questionnaire

Epstein, J. L., & McPartland, J. M. (1978). *The Quality of School Life Scale*. Boston, MA: Houghton Mifflin.

Safety

Feelings of Safety Items

Schwab-Stone, M., Ayers, T., Kasprow, W., Voyce, C., Barone, C., Shriver, T., & Weissberg, R. (1995). No safe haven: A study of violence exposure in an urban community. *Journal of the American Academy of Child and Adolescent Psychiatry*, 34, 1343-1352.

Satisfaction

Client Satisfaction Questionnaire-8

Larsen, D. L., Atkisson, C. C., Hargreaves, W. A., & Nguyen, T. D. (1979). Assessment of client/patient satisfaction: Development of a general scale. *Evaluation and Program Planning*, 2, 139-150.

Consumer Satisfaction Questionnaire (author unknown)

Satisfaction Scales

Brannan, A. M., Sonnichsen, S. E., & Heflinger, C. A. (1996). Measuring satisfaction with children's mental health services: Validity and reliability of the satisfaction scales. *Evaluation and Program Planning*, 19, 131-141.

Youth Satisfaction Questionnaire

Stuentzner-Gibson, D., Koren, P. E., & DeChillo, N. (1995). The Youth Satisfaction Questionnaire: What kids think of services. *Families in Society*, 76, 616-624.

Self-Esteem

Canadian Self-Esteem Inventory for Children

Battle, J. (1976). Test-retest reliability of the Canadian Self-Esteem Inventory for Children. *Psychological Reports*, 38, 1343-1345.

Coopersmith Self-Esteem Inventories

Coopersmith, S. (1967). *The antecedents of self-esteem*. Palo Alto, CA: Consulting Psychologists Press.

Coopersmith, S. (1982). *Self-esteem inventories*. Palo Alto, CA: Consulting Psychologists Press.

Culture-Free Self-Esteem Inventories II

Battle, J. (1992). *Culture-Free Self-Esteem Inventories: Examiner's manual* (2nd ed.). Austin, TX: Pro-Ed.

Global Self-Worth scale

Harter, S. (1982). The Perceived Competence Scale for Children. *Child Development*, 53, 87-97.

Lipsitt Self-Concept Scale

Lipsitt, L. P. (1958). A self-concept scale for children and its relationship to the children's form of the Manifest Anxiety Scale. *Child Development*, 29, 463-472.

Multidimensional Self Concept Scale

Bracken, B. A., & Howell, K. K. (1991). Multidimensional self-concept validation: A three-instrument investigation. *Journal of Psychoeducational Assessment*, 9, 319-328.

Offer Self-Image Questionnaire for Adolescents -Revised

Offer, D., Ostrov, E., & Howard, K. I. (1982). Family perceptions of adolescent self-image. *Journal of Youth and Adolescence*, 12, 117-135.

Perceived Competence Scale for Children

Harter, S. (1982). The Perceived Competence Scale for Children. *Child Development*, 53, 87-97.

Perceived Self-Worth Questionnaire (author unknown)

Personal Attribute Inventory for Children revised by Parish and Rankin (1982) and renamed the Nonsexist Personal Attribute Inventory for Children (NPAIC).

Parish, T., & Taylor, J. (1978a). A further report on the validity and reliability of the Personal Attribute Inventory for Children as self-concept scale. *Educational and Psychological Measurement*, 38, 1225-1228.

Parish, T., & Taylor, J. (1978b). The Personal Attribute Inventory for Children: A report on its validity and reliability as a self-concept scale. *Educational and Psychological Measurement*, 38(2), 565-569.

Parish, T., & Rankin, C. I. (1982). The Nonsexist Personal Attribute Inventory for Children: A report on its validity and reliability as a self-concept scale. *Educational and Psychological Measurement*, 42, 339-343.

Piers-Harris Self-Concept Scale; also called: Children's Self-Concept Scale and The Way I Feel About Myself

Piers, E. V. (1984). *Piers-Harris Children's Self-Concept Scale: Revised manual*. Los Angeles: Western Psychological Services.

Piers, E., & Harris, D. (1969). *The Piers-Harris Children's Self-Concept Scale*. Nashville: Counselor Recordings and Tests.

Rosenberg Self-Esteem Scale

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Self-Appraisal Inventory

Narikawa, O., & Frith, S. (1972). *Measures of self-concept*. Los Angeles: Instructional Objectives Exchange.

Self-Description Inventory

Mboya, M. (1993). Development and construct validity of self-description inventory for African adolescents. *Psychological Reports*, 72, 183-191.

Self-Description Questionnaire I

Marsh, H. W., Barnes, J., Cairnes, L., & Tidman, M. (1984). Self-Description Questionnaire: Age and sex effects in the structure and level of self-concept for preadolescent children. *Journal of Educational Psychology*, 76, 940-956.

Self-Description Questionnaire II

Marsh, H. W. (1992). *Self-Description Questionnaire-II manual*. University of Western Sydney, Cambellton, Australia.

Self-Description Questionnaire III

Marsh, H. W., Richards, G. E., & Barnes, J. (1986). Multidimensional self-concepts: The effect of participation in an outward bound program. *Journal of Personality and Social Psychology,* 50, 195-204.

Self-Esteem Index

Brown, L., & Alexander, J. (1991). *Self-Esteem Index examiner's manual*. Austin, TX: Pro-Ed.

Self-Perceived Role Competence

Beiser, M., Lancee, W., Gotowiec, A., Sack, W., & Redshirt, R. (1993). Measuring self-perceived role competence among First Nations and non-native children. *Canadian Journal of Psychiatry*, 38, 412-419.

Self-Perception Profile for Adolescents

Harter, S. (1988). *Manual for the Self-Perception Profile for Adolescents*. Denver, CO: University of Denver.

Self-Perception Profile for Children

Harter, S. (1988). *Manual for the Self-Perception Profile for Children*. Denver, CO: University of Denver.

Tennessee Self-Concept Scale

Roid, G. H., & Fitts, W. (1988). *Tennessee Self-Concept Scale: Revised manual*. Los Angeles: Western Psychological Services.

What I Think About Myself

Mathewson, R. H., & Orton, J. W. (1963). Vocational imagery and vocational maturity of high school students. *Journal of Counseling Psychology*, 10, 384-388.

Social Support

Family, Friends, and Self Assessment Scales

Simpson, D. D., & McBride, A. A. (1992). Family, Friends, and Self (FFS) Assessment Scales for Mexican American youth. *Hispanic Journal of Behavioral Sciences*,14, 327-340.

Perceived Social Support Scale from Family (PSS-FA) and Perceived Social Support Scale from Friends

Procidiano, M. E., & Heller, K. (1983). Measures of perceived social support from friends and from family: Three validation studies. *American Journal of Community Psychology*, 11, 1-24.

Perceived Social Support from School Personnel

DuBois, D. L., Felner, R. D., Brand, S., Adan, A. M., & Evans, E. G. (1992). A prospective study of life stress, social support, and adaptation in early adolescence. *Child Development*, 63, 542-557.

Personal History Inventory for Children

Parish, T. S., & Wigle, S. E. (1985). Discerning functionality of children's support systems through use of the Personal History Inventory for Children. *Psychological Reports*, 57, 32-34.

Social Support Appraisals scale – Revised

Dubow, E. F., & Ullman, D. G. (1991). Assessing social support in elementary school children: The survey of children's social support. *Journal of Clinical Child Psychology*, 18, 52-64.

Social Support Scale for Children and Adolescents

Harter, S. (1985). *Manual for the Social Support Scale for Children*. Denver, CO: University of Denver.

Supportive Parenting Scale

Simons, R. L., Lorenz, F. O., Conger, R. D., & Wu, C. (1992). Support from the spouse as a mediator and moderator of the disruptive influence of economic strain on parenting. *Child Development*, 63, 1282-1301.

Stressful Events

Academic Pressure Scale for Adolescents

West, C. K., & Wood, E. S. (1970). Academic pressures on public school students. *Educational Leadership*, 3, 585-589.

Adolescent Perceived Events Scale

Compas, B. E., Davis, G. E., Forsythe, C. J., & Wagner, B. M. (1987). Assessment of major and daily stressful events during adolescence: The Adolescent Perceived Events Scale. *Journal of Consulting and Clinical Psychology*, 55, 534-541.

Children’s Own Perceptions and Experiences of Stressors

Colton, M. (1989). Foster and residential children's perceptions of their social environments. *British Journal of Social Work*, 19, 217-223.

Daily Hassles Questionnaire

Rowlison, R. T., & Felner, R. D. (1988). Major life events, hassles, and adaptation in adolescence: Confounding in the conceptualization and measurement of life stress and adjustment revisited. *Journal of Personality and Social Psychology*, 55, 432-444.

Family Inventory of Life Events

McCubbin, H. I., & Patterson, J. M. (1983). The family stress process: The Double ABCX Model of Adjustment and Adaptation. *Marriage and Family Review*, 6, 7-37.

Hassles Scale for Children

Parfenoff, H., & Jose, P. E. (1989). *Measuring daily hassles in children*. Paper presented at the Biennial meeting of the Society for Research in Child Development, Kansas City, Missouri.

Life Events Checklist

Johnson, J. H., & McCutcheon, S. M. (1980). Assessing life stress in older children and adolescents: Preliminary findings with the Life Events Checklist. In I. G. Sarason, & C. D. Spielberger (Eds.), *Stress and anxiety* (Vol. 7, pp. 111-125). Washington, DC: Hemisphere.

Life Events Scale for Children

Coddington, R. (1972). The significance of life events as etiologic factors in the diseases of children: A study of a normal population. *Journal of Psychosomatic Research*, 16, 205-403.

Sandler and Block’s (1979) Modification of the Coddington Life Events Scale for Children

Sandler, I. N., & Block, M. (1979). Life stress and maladaptation of children. *American Journal of Community Psychology*, 57, 316-331.

Therapeutic Process

Adolescent Working Alliance Inventory

Linscott, J., DiGiuseppe, R., & Jilton, R. (1993). *A measure of therapeutic alliance in adolescent psychotherapy*. Paper presented at the 101st Annual Convention of the American Psychological Association, Toronto, Ontario, Canada.

Barriers-to-Treatment Participation Scale

Kazdin, A. E., Holland, L., Crowley, M., & Breton, S. (1997). Barriers to Treatment Participation Scale: Evaluation and validation in the context of child outpatient treatment. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 38, 1051-1062.

Family Involvement Items

Baker, B., Blacher, J., & Pfeiffer, S. (1993). Family involvement in residential treatment of children with psychiatric disorder and mental retardation. *Hospital and Community Psychiatry*, 44, 561-566.

Loyola Child Psychotherapy Process Scales

Estrada, A. U., & Russell, R. L. (1994). *The development of the Loyola Child Psychotherapy Process Scales*. Unpublished Manuscript.

Motivation to Change Questionnaire

Phares, V., & Danforth, J. (1994). Adolescents', Parents', and Teachers' Distress over Adolescents' Behavior. *Journal of Abnormal Child Psychology*, 22, 721-732.

Penn Helping Alliance Questionnaire; client and therapist versions

Alexander, L. B., & Luborsky, L. (1986). The Penn Helping Alliance Scales. In L. S. Greenberg, & W. M. Pinsof (Eds.), *The psychotherapeutic process: A research handbook* (Vol. 18, pp. 325-366). New York: Guilford Press.

Psychotherapy Process Inventory (modified for use with children)

Gorin, S. (1993). The prediction of child psychotherapy outcome: Factors specific to treatment. *Psychotherapy*, 30, 152-158.

Smith-Acuna, Durlak, & Kaspar (1991) (unnamed scales)

Smith-Acuna, S., Durlak, J. A., & Kaspar, C. J. (1991). Development of a child psychotherapy process measure. *Journal of Clinical Child Psychology*, 20, 126-131.

Therapeutic Alliance Scales for Children

Shirk, S., & Saiz, C. (1992). Clinical, empirical, and developmental perspectives on the therapeutic relationship in child psychotherapy. *Development and Psychopathology*, 4, 713-728.

Therapy Process Checklist

Weersing, V. R. (1996). Development and Validation of the Therapy Process Checklist. Unpublished doctoral dissertation, University of California at Los Angeles.

Treatment Acceptability Questionnaire

Hunsley, J. (1992). Development of the Treatment Acceptability Questionnaire. *Journal of Psychopathology and Behavioral Assessment*, 14, 55-64.

Treatment Evaluation Inventory (TEI) and TEI-Short Form

Kazdin, A. E. (1980). Acceptability of alternative treatments for child deviant behavior. *Journal of Applied Behavior Analysis*, 13, 259-273.

Kelley, M. L., Heffer, R. W., Gresham, F. M., & Elliot, S. N. (1989). Development of a modified Treatment Evaluation Inventory. *Journal of Psychopathology and Behavioral Assessment*, 11, 235-247.

CHAPTER VII
APPENDICES

APPENDIX 1

A GLOSSARY OF TERMS USED IN THIS REVIEW

Acuity: The degree to which the patient is a danger to self or others, by virtue of suicidality, self-injuriousness, impulsivity, aggressiveness, disorganised thinking, lack of awareness of environmental dangers, or failure to attend to vital bodily needs.

Carer: Somebody who has the responsibility of caring for somebody else, in this case a child or adolescent with mental illness.

Chronicity: The relative time during which the patient has suffered (mental health) symptoms and functional impairment, with implications for reversibility.

Consumer: Somebody who uses a product or receives (mental health) services.

Level of care: A specified type of (mental health) care varying in intensity (and cost) in accordance with the patient's acuity, severity, and type of problem. Levels of care can range in intensity from outpatient to hospitalization.

Measure: The means by which the proportions of a specified characteristic of something can be ascertained, relative to a standard. An *outcome measure* is an instrument designed to estimate the extent to which a planned outcome has been (or is being) attained.

Mediating factor: The means or process by which change occurs between input and outcome.

Moderating factor: Antecedent or concurrent events or factors that impede or facilitate process (mediating factors) and affect the extent to which a planned outcome can be attained.

Outcome: The practical consequence, effect, or issue of an action. A *mental health outcome* is the effect of mental health services on the health status of individual (or groups of) mental health patients and their families. A *mental health outcome domain* is a specified type or aspect of mental health outcome.

Process: A series of connected actions aimed to produce an outcome or product.

Provider: A clinician who delivers (mental) health services.

Severity: The frequency, intensity and diversity of a patient's symptomatology together with the extent of functional impairment.

System: A complex whole composed of interrelated parts. A *Mental Health System* is an identified organisation, usually in the public sector, the purpose of which is to treat and prevent mental illness. Mental health systems coordinate such components as inpatient units, partial hospital units, outpatient services, and outreach services and are linked to other organisations that have related aims.

Treatment: The process by which a clinician provides services to patients (and their families) with the aim of improving their (mental) health. A *treatment modality* is a particular form of treatment prescribed in accordance with the patient's diagnosis. A *treatment protocol* involves a formalised set of steps to be followed by all clinicians who deliver that treatment.

APPENDIX 2

SEMI-STRUCTURED QUESTIONNAIRE

Introduction to Interviewees

As you are aware the Commonwealth Department of Health and Family Services has contracted with the University of Queensland, Division of Child and Adolescent Psychiatry to undertake a review of existing treatment outcome measures in child and adolescent mental health.

The aim is to develop a Treatment Outcome Measurement System either by adapting existing measures or by designing a new system and measures.

Why we need your assistance

We are seeking input from experts in child and youth mental health such as yourself, to ascertain the status of outcome measures for children and adolescents.

1. We would like to know what measures are being used and the adequacy of these measures.
2. We would like your opinion as to whether a comprehensive outcome measurement system is already in existence and if not, if such a system should be developed.
3. We wish to explore with you what this system should look like if it were to be developed, and the issues and challenges of design and implementation.

1. Respondent's Background

May I quickly check these details with you.

Name/Title: ....

Address: ....

....

Current position: ....

(a) Could you outline any experience you have had in the area of Outcome Measurement in child and youth mental health?

(b) Have you ever treated children and/or young people with mental health problems? (If yes: *can you tell me briefly about it?*)

(c) Have you used outcome measurements in any of the settings in which you have worked?
(*Prompts:* *Which?*
How did you use it?
Was it helpful?
In what ways?
Were there ways you think the measurement or use could have been improved?)

(d) Have you undertaken any research in outcome measurement in child and/or youth mental health? (If yes: *Can you tell me briefly about it?*)

(e) Do you have any particular areas of interest in outcome measurement?

2. Broad view of outcome measurement systems

(a) How would you define an outcome measurement system?

(b) What do you think are the ideal criteria for such a system?

(*Prompts:* *feasibility*
comprehensiveness
potential for improving clinical effectiveness
multiple outcome levels
psychometric soundness
developmental sensitivity
cultural sensitivity)

It has been argued that an outcome system should assess the different outcome domains that stakeholders consider important. An assessment of single domains will miss changes that occur in other domains.

(c) How comprehensive do you think this system should be?

(d) What domains or areas of measurement should such a system encompass?

Prompts: *Should it provide information about the following?*

	Yes	No
• diagnostic data	☐	☐
• general symptomatology	☐	☐
• adaptive functioning	☐	☐
• goals of treatment	☐	☐
• child and family environment	☐	☐
• burden of care-quality of life	☐	☐
• consumer satisfaction	☐	☐
• service and system	☐	☐
• motivation and readiness to change	☐	☐
• quality of therapeutic alliance	☐	☐
• treatment modality	☐	☐

I'd like to further explore the question of domains or areas of measurement by going through a list of outcomes domains. Please tell me whether you think an outcome measurement system should provide information for each of these domains.

YOUTH-ORIENTED CATEGORIES

	Yes	No
1. Global score for the youth's problems (e.g., global severity score)	☐	☐
2. Symptom scores for specific disorders/problem areas for youths (e.g., depression, anxiety, drug/ alcohol use, eating disorders, suicidal ideation)	☐	☐
3. Areas specific to youth's level of functioning (e.g., school, peers, family, community)	☐	☐
4. Youth's global level of functioning score (e.g., functioning score aggregated across domains: school, work, peer groups, family, community)	☐	☐
5. Youth skills and strengths (e.g., social skills, hobbies, and other positive activities)	☐	☐
6. Youth's physical health and current medications	☐	☐
7. History of child maltreatment (i.e., physical abuse, sexual abuse, emotional abuse, and neglect)	☐	☐
8. Past and present youth stressors (e.g., death of a parent, poor grades, poor relations with family and friends)	☐	☐
9. Youth's motivation to change	☐	☐
10. Youth's reasons for seeking therapeutic services	☐	☐
11. Youth ‘s attitude about seeking mental health services (e.g., optimistic about potential benefits of services, dislike of mental health professionals)	☐	☐
12. Youth's past utilisation of mental health services, drug/alcohol treatment, and medical services	☐	☐

Do you think an outcome measurement system should provide information for each of these domains?

PARENT / FAMILY-ORIENTED CATEGORIES

	Yes	No
13. Parent's physical and mental health, behavioural symptoms and history with psychotropic medications	☐	☐
14. Parent's reasons for seeking services for youth	☐	☐
15. Parental attitudes about seeking therapeutic services for youth (e.g., optimistic about potential benefits of services, dislike for mental health professionals)	☐	☐
16. Parent's stress and strain of caring for youth with mental health problems	☐	☐
17. Parental expectations for youth with respect to current and future personal and career goals	☐	☐
18. Quality of parent-youth relationship	☐	☐
19. Family functioning (e.g., family relations, conflict, communication among family members)	☐	☐
20. Past and present family stress (e.g., job loss, family member married, childbirth or parent died)	☐	☐
21. Concrete family resources to meet family needs (e.g., income, health insurance, housing, food)	☐	☐
22. Barriers to care (e.g., financial, transportation, school schedule, parent/teen's work schedule)	☐	☐

Do you think an outcome measurement system should provide information for each of these domains?

THERAPY-RELATED CATEGORIES

	Yes	No
23. Youth involvement in developing treatment plan	☐	☐
24. Parent involvement in developing treatment plan	☐	☐
25. Parent's compliance with treatment plan (e.g., follow through with disciplinary actions, assist youth in meeting weekly therapy goals)	☐	☐
26. Youth's compliance with treatment plan (e.g., completion of therapy "homework" assignments)	☐	☐
27. Therapeutic alliance (e.g., parent or youth's perception of their relationship with the clinician)	☐	☐
28. Satisfaction with youth's therapist, clinic, and service system.	☐	☐
29. Narrative (open ended) responses from parents or youths on various topics including: medications, goal attainment, or progress (e.g., How effective are the medications?, Do you think you are achieving your goals from therapy?, Overall, are your child's emotional and/or behavioural problems improving, staying the same or getting worse?)	☐	☐

30. Can you think of any other domain not included in the list above?

(e) How often should information be collected?
(Probe if not volunteered)

Should information be collected –

		Yes	No
a.	At intake	☐	☐
b.	Concurrent with treatment	☐	☐
c.	At termination of treatment	☐	☐
d.	6 months after starting treatment	☐	☐
e.	12 months after starting treatment	☐	☐
f.	Longer than 12 months after starting treatment	☐	☐

(f) Who should be the sources of information?
(Probe if not volunteered)

Should it include?

		Yes	No
a.	Child (2-12)	☐	☐
b.	Youth (over 12)	☐	☐
c.	Caregiver or parent	☐	☐
d.	Teacher	☐	☐
e.	Mental health clinician	☐	☐
f.	Other – who?	☐	☐

(g) How would you gather this information?
Prompts:
questionnaires, computerised questionnaires, face-to-face interviews, mailed out questionnaires.

(h) Do you think a measurement system can help clinicians treat clients?
If yes: *In what way*?

(i) Should a measurement system be used by organizations to measure the quality of their services?

(j) Should standard measures be used throughout Australia or should each organisation use whatever measure they wish?
If the latter, should there be standards that organisations have to meet? If so, what standards, how rigorous?

(k) Should a measurement system be used by the federal government to evaluate cost effectiveness in treatment?

(l) Are you aware of any outcome measurement system that should be used in child mental health, or should we build one from scratch?

Suppose that such a system included the areas of measurement we discussed earlier, and that data were collected from the youth, caregiver or parent and the clinician, at intake, briefly concurrent with treatment (every session or at regular intervals while the youth was in treatment) and 6 and 12 months after starting treatment, and the same measurement instruments were used throughout Australia.

(m) How important would such a system be to child and youth mental health?

(n) How much support would there be for such a system throughout all areas of child and youth mental health?

(o) Would you adopt or encourage adoption of such a system in your area? Why or Why not?

(p) What investment and commitment would researchers, administrators, and clinicians be prepared to devote to a system like this?

(q) Are there changes that you would make to the system we have described earlier that would make it acceptable to you? To others?

3. Feasibility

If a comprehensive system of outcome measurement is to be developed, feasibility will be crucial. I would like your views as to how an outcome system could be made applicable, acceptable, and practicable.

1. How much support within child and youth mental health do you believe there will be for such a system?

2. What investment and commitment would mental health clinicians be prepared to make to such a system?

3. What investment and commitment would parents, caregivers and clients be prepared to make to such a system?

4. Do you think mental health staff will be receptive to using data from an outcome measurement system?

5. Do you think mental health clinicians will be able to use data from an outcome measurement system?
   If not, what would need to be done to make such data useful?

6. How much time and resources can child and youth mental health services reasonably devote to collecting information for outcome measurement?

7. How much time do you think clinicians can devote to collecting data, per session?
   ---------mins?

8. How much time can clinicians devote to reviewing feedback data per session?

9. How much time do you think clients, parents or caregivers and teachers can devote to providing data, per session? ---------mins?

10. Would clinics need to have a computerized system to collect data?
    To analyze data? To distribute feedback data to clinicians?

11. Should clinicians be trained to make use of outcome data? How should they be trained?

Implementation

12. What risks do you anticipate in the implementation of such a system?
    (Harm to patients, medical/legal? How will it be used by whom, for what)?

13. Do you think that such a system could have adverse effects?
    (If Yes: can you please elaborate?)

14. Do you anticipate barriers to the implementation of an outcome measurement system?
    (If Yes: *What are they?*)

15. In terms of barriers, do you think there will be significant clinician resistance to a treatment outcome measurement system?

16. Do you think there will be significant resistance from clients and parents or caregivers to such a system?

17. Do you think there will be significant resistance from directors or managers of mental health services or clinics?

18. What could be done to avert or mitigate resistance?

APPENDIX 3

FORMAT FOR FOCUS GROUP DISCUSSION

Focus Group Discussion Framework for CYMHS Clinicians

Introduction

- Provide brief background to outcomes consultancy
- Outline purpose of this discussion with emphasis on improving child and youth mental health services.
- Outline ethical standards:

  Recording of discussion

  Participants are not identifiable

  Participants may leave at any time

  Participants will be able to access the final report
- Provide brief explanation of focus groups including:

  Rules setting

  Nature of discussion (focussed rather than general)

  Outline what is sought from the group

Issues for Guided Discussion

1. Evaluation in child and youth mental health services (Briefly present and discuss a hypothetical consumer outcome measurement model)

- Have participants ever experienced outcome measurement in child and youth mental health?
- Exploration of previous experiences in work settings or in research.
- If participants have experienced outcome measurement, identify instruments or systems, utilisation and perceived usefulness. Examine whether the measurement or use of it could have been improved.
- Based on participants' experience as a CYMHS clinician, what aspects of CYMHS could most benefit from an evaluation system?

2. Broad view of outcome measurement systems

- How would participants define an outcome measurement system? What do participants consider are the ideal criteria/characteristics for such a system? (Prompts: feasibility, comprehensiveness, potential for improving clinical effectiveness, multiple outcome levels, psychometric soundness, developmental sensitivity, cultural sensitivity.
- How comprehensive should an outcome measurement system be?
- What are the different outcome domains that stakeholders consider important? (Prompts: Severity of child or adolescents symptoms, dangerousness or acuity of child or adolescent's symptoms, level of adaptive functioning, goals of treatment)

3. Data Collection

- Frequency of data collection
- Sources of information

4. Importance of a measurement system to child and adolescent mental health (Refer to the original hypothetical consumer outcome model and the group's broad view of outcome measurement systems)

- How important would such a system be to child and youth mental health?
- Do participants think a good measurement system can help clinicians treat clients?
- Should a measurement system be used by organizations such as CYMHS to measure the quality of their services?
- Should a good measurement system be used by Government to evaluate their investment of funds in treatment?
- What percentage of all money spent on child and youth mental health should be spent on maintaining an ongoing outcome measurement system?
- Should standard measures be used throughout Australia or should each organisation use whatever measure they wish?
- If standard measures should not be adopted, should there be standards those organizations have to meet? If so, what standards should be met and how rigorously?

5. Build, Borrow or Buy

- Are you aware of any outcome measurement system that should be used in child and adolescent mental health, or should one be built from scratch?

6. Support and Feasibility (Outline conceptual model)

- How important would such a system be to child and youth mental health?
- How much support would there be for such a system throughout all areas of child and youth mental health particularly CYMHS?
- Would participants adopt or encourage adoption of such a system within CYMHS?
- What investment and commitment would CYMHS staff be prepared to devote to a system like this?
- Are there changes that participants would make to the system we have described earlier that would make it acceptable to them? To other CYMHS staff?
- What investment and commitment would parents, caregivers and clients be prepared to make to such a system?
- Do participants think CYMHS staff will be receptive to using data from an outcome measurement system?
- Do participants think CYMHS case managers will be able to use data from an outcome measurement system?
- If not, what would need to be done to make such data useful?
- How much time and resources can child and youth mental health services such as CYMHS reasonably devote to collecting information for outcome measurement?
- How much time do participants think CYMHS case managers can devote to collecting and reviewing feedback data, per session? ————mins?
- How much time do participants think clients, parents or caregivers and teachers can devote to providing data, per session? ————mins?
- Would clinics need to have a computerized system to collect data?
  To analyze data? To distribute feedback data to clinicians?
- Should clinicians be trained to make use of outcome data? How should they be trained?

7. Potential Challenges and Barriers

- What risks do participants anticipate in the implementation of such a system? (Harm to patients, medical/legal? How will it be used, by whom, for what reasons, what adverse effects could be anticipated)?
- Do participants anticipate barriers to the implementation of an outcome measurement system within and/or outside CYMHS?

- In terms of barriers, do you think there will be significant clinician resistance to a treatment outcome measurement system?
- Do participants think there will be significant resistance from clients and parents or caregivers to such a system?
- Do participants think there will be significant resistance from directors or managers of mental health services or clinics?
- What could be done to avert or mitigate resistance?

Focus Group Discussion Framework for Adolescents Accessing a Child and Youth Mental health Service (CYMHS)

Introduction

- Provide brief background to outcomes consultancy
- Outline purpose of this discussion with emphasis on improving child and youth mental health services.
- Outline ethical standards:

  Recording of discussion

  Participants are not identifiable

  Participants may leave at any time

  Participants will be able to access the final report
- Provide brief explanation of focus groups including:

  Rule setting

  Nature of discussion (focussed rather than general)

  Outline what is sought from the group

Issues for Guided Discussion

1. Hypothetical model of outcome measurement system (Briefly present and discuss a hypothetical consumer outcome measurement model)

- Ascertain participants' previous experience with evaluation either at work/study or as consumers of services.
- Compare participants' experiences to the hypothetical model
- Develop the hypothetical model based on participants' experience of an event, such as a hospital stay, where they were a service consumer.

2. Evaluation in child and youth mental health services

- Refer to hypothetical outcomes model, explore whether participants have experienced a similar system in child and youth mental health.
  (Prompts: have participants filled out questionnaires, at which point during the service, what do participants think was/were being measured?)
- If participants have experienced a similar system, was this helpful? How was it helpful? Explore whether participants think the measurement or the use of it could have been improved?
- Based on participants' experiences as consumers of CYMHS services, what aspects of these services do they think could most benefit from an evaluation system?

3. Broad view of outcome measurement systems

- How would participants define a 'good' outcome measurement system?
- What do participants consider important criteria or standards of an outcome measurement system (Prompts: feasibility, comprehensiveness, potential for improving clinical effectiveness, multiple outcome levels, psychometric soundness, developmental sensitivity, cultural sensitivity.)
- How comprehensive should an outcome measurement system be?
- What are the different outcome domains that stakeholders consider important? (Prompts: Severity of child or adolescent's symptoms, dangerousness or acuity of child or adolescent's symptoms, level of adaptive functioning, goals of treatment).
- What are the indicators of 'good' outcomes? Good for whom?
- What are the indicators of 'bad' outcomes? Bad for whom?
- What else should this system measure? (E.g. Processes; Service provision)

4. Data Collection

- Frequency of data collection
- Sources of information

5. Importance of a measurement system to child and adolescent mental health (Refer to the hypothetical model and the group's broad view of the consumer outcome measurement systems)

- How important would such a system be to child and youth mental health?
- Do participants think a good measurement system can help clinicians treat clients?

- Should a good measurement system be used by an organisation such as CYMHS to measure the quality of the service provided to clients?
- Should a good measurement system be used by Government to evaluate their investment of funds in treatment
- What percentage of all money spent on child and youth mental health should be spent on maintaining an ongoing outcome measurement system?
- Should standard measures be used throughout Australia?
- If standard measures should not be used, should there be standards that organizations have to meet? If so, what should these standards be like? How rigorous?

6. Feasibility and Support

- Would adolescents who use CYMHS be willing to support an outcome measurement system?
- How would participants and parents like to provide information? (Prompts: questionnaires, face-to-face interviews, mailed interviews, computer assisted questionnaires).
- Who should provide information? (Prompts: Parents, carers, adolescents, children, teachers)
- How much time do participants think clients, parents and teachers would be willing to devote to providing information, such as filling out questionnaires, per session?
- Would CYMHS clinicians be able to use data from an outcome measurement system?
- Should clinicians be trained to make use of outcome data? How should they be trained?

7. Potential Challenges and Barriers

- Do participants anticipate barriers to the implementation of an outcome measurement system in clinics such as CYMHS?
- Do participants anticipate clinician resistance to a treatment outcome measurement system?
- Do participants anticipate client or parental and carer resistance to a treatment outcome measurement system?
- What are the anticipated risks or adverse effects in the implementation of such a system? (To clinicians, clients, the service)
- How could potential challenges and barriers be addressed?

Focus Group Discussion Framework for Parents and Carers of Children and Young People Accessing a Child and Youth Mental Health Service (CYMHS)

Introduction

- Provide brief background to the outcomes consultancy
- Outline purpose of this discussion with emphasis on improving child and youth mental health services.
- Outline ethical standards:

  Recording of discussion

  Participants are not identifiable

  Participants may leave at any time

  Participants will be able to access the final report
- Provide brief explanation of focus groups including:

  Rule setting

  Nature of discussion (focussed rather than general)

  Outline what is sought from the group

Issues for Guided Discussion

1. Hypothetical model of outcome measurement system (Briefly present and discuss a hypothetical consumer outcome measurement model)

- Ascertain participants' previous experience with evaluation either at work/study or as consumers of services.

  Compare participants' experiences to hypothetical model

  Develop the hypothetical model based on participants' experiences of an event where they were a service consumer, such as giving birth in a hospital setting.

2. Evaluation in child and youth mental health services

- Refer to hypothetical outcomes model, explore whether participants have experienced a similar system in child and youth mental health.
  (Prompts: have participants filled out questionnaires, at which point during the service, what do participants think was/were being measured?)

- If participants have experienced a similar system, was this helpful? How was it helpful? Explore whether participants think the measurement or the use of it could have been improved?
- Based on participants' experiences as consumers of CYMHS services, what aspects of these services do they think could most benefit from an evaluation system?

3. Broad view of outcome measurement systems

- How would participants define a 'good' outcome measurement system?
- What do participants consider important criteria or standards of an outcome measurement system? (Prompts: feasibility, comprehensiveness, potential for improving clinical effectiveness, multiple outcome levels, psychometric soundness, developmental sensitivity, cultural sensitivity).
- How comprehensive should an outcome measurement system be?
- What are the different outcome domains that stakeholders consider important? (Prompts: Severity of the child or adolescent's symptoms, dangerousness or acuity of child or adolescent's symptoms, level of adaptive functioning, goals of treatment).
- What are the indicators of 'good' outcomes? Good for whom?
- What are the indicators of 'bad' outcomes? Bad for whom?
- What else should this system measure? (E.g. Processes; Service provision)

4. Data collection

- Frequency of data collection
- Sources of information

5. Importance of a measurement system to child and adolescent mental health (Refer to the hypothetical model and the group's broad view of the consumer outcome measurement systems)

- How important would such a system be to child and youth mental health?
- Do participants think a good measurement system can help clinicians treat clients?
- Should a good measurement system be used by an organisation such as CYMHS to measure the quality of the service provided to clients?
- Should a good measurement system be used by Government to evaluate their investment of funds in treatment?

- What percentage of all money spent on child and youth mental health should be spent on maintaining an ongoing outcome measurement system?
- Should standard measures be used throughout Australia?
- If standard measures should not be used, should there be standards that organizations have to meet? If so, what should these standards be like? How rigorous?

6. Feasibility and support

- Would parents, clients and other carers be willing to support an outcome measurement system?
- How would parents and clients like to provide information?(Prompts: questionnaires, face-to-face interviews, mailed interviews, computer assisted questionnaires)
- Who should provide information? (Prompts: Parents, carers, adolescents, children, teachers.)
- How much time do participants think clients, parents and teachers would be willing to devote to providing information, such as filling out questionnaires, per session?
- Would CYMHS clinicians be able to use data from an outcome measurement system?
- Should clinicians be trained to make use of outcome data? How should they be trained?

7. Potential challenges and barriers

- Do participants anticipate barriers to the implementation of an outcome measurement system in clinics such as CYMHS?
- Do participants anticipate clinician resistance to a treatment outcome measurement system?
- Do participants anticipate client or parental and carer resistance to a treatment outcome measurement system?
- What are the anticipated risks or adverse effects in the implementation of such a system? (To clinicians, clients, the service)
- How could potential challenges and barriers be addressed?

APPENDIX 4

CONSUMER OUTCOMES QUESTIONNAIRE

This questionnaire seeks your general opinion about ways of measuring the results of treatment for children and adolescents who use mental health services. Please answer each of the following questions by choosing a response that best suits your views. There are five answers to choose from. These are Not at all, A little, Somewhat, Very, and Extremely. Please tick the number that best matches your answer to each question.

	Not at all	A little	Somewhat	Very	Extremely
1. How important is it that mental health services know how helpful or unhelpful their services are/have been for children and adolescents who use their services?	1	2	3	4	5
2. When finding out how helpful or unhelpful our services are/or have been for you and your child, how important is it that we collect information about the following?					
Your child's problems (eg. feeling low, depressed or anxious)	1	2	3	4	5
Details about your child and family environment (eg home, school, neighbourhood)	1	2	3	4	5
How your child is doing at home, at school and with friends	1	2	3	4	5
What you or your child and the case manager hope to achieve with treatment	1	2	3	4	5
Your family's quality of life	1	2	3	4	5
Your satisfaction with services	1	2	3	4	5
How you and your child get on with your child's case manager	1	2	3	4	5
Other? ____________________ _________________________					

		Not at all	A little	Somewhat	Very	Extremely
3.	Would you like to use the following to provide information about yourself or your child?					
	A paper questionnaire which you fill out in the clinic	1	2	3	4	5
	A questionnaire which you fill out using a computer in the clinic	1	2	3	4	5
	A questionnaire sent to your home	1	2	3	4	5
	A face to face interview in the clinic with someone other than your child's case manager	1	2	3	4	5
	Other? ____________________ __________________________					
4.	When finding out how helpful or unhelpful our services are/or have been for you and your child, how important is it that we get information from the following people?					
	The client aged between 2 to 6 years	1	2	3	4	5
	The client aged between 7 to 11 years	1	2	3	4	5
	The client aged 12 years or over	1	2	3	4	5
	The parent or caregiver	1	2	3	4	5
	Teacher(s)	1	2	3	4	5
	The case manager	1	2	3	4	5
	Other? (please specify) __________________________					

	Not at all	A little	Somewhat	Very	Extremely
5. When finding out how helpful our services are/have been for you and your child, how important is it that we obtain information from you at these times?					
At your child's first interview?	1	2	3	4	5
During your child's treatment?	1	2	3	4	5
At the end of treatment?	1	2	3	4	5
Three months after finishing treatment?	1	2	3	4	5
Six months after finishing treatment?	1	2	3	4	5
Twelve months after finishing treatment?	1	2	3	4	5
Other? (please specify) ______________________					
6. How important is it that the information collected from you is used for the following?					
To provide information to your case manager about your child's progress	1	2	3	4	5
To provide information to you – the parent and/or caregiver about your child's progress	1	2	3	4	5
To provide information to the child or adolescent about their own progress	1	2	3	4	5
To allow organisations to assess how good their services are	1	2	3	4	5
To allow governments to make decisions about funding	1	2	3	4	5

7. Would you like to make any other comments? If so, please write in the space below.

____

____

____

Thank you for completing this questionnaire.

APPENDIX 5

LIST OF MEASURES REVIEWED

Coping Skills (11)

Adolescent-Coping Orientation for Problem Experiences Scale (Patterson & McCubbin, 1987)

Adolescent Coping Scale (Frydenberg & Lewis, 1993)

Analysis of Coping Style (Boyd & Johnson, 1981)

Coping Inventory for Adolescents (Fanshawe & Burnett, 1991)

Coping Responses Inventory (Moos, 1997)

Coping Strategies Inventory (Tobin, Holroyd, & Reynolds, 1984)

Kidcope (Spirito, Stark, & Williams, 1988)

Problem Inventory for Adolescent Girls (Gaffney, 1984)

Social Coping Questionnaire for Gifted Students (Swiatek, 1995)

Stress Response Scale (Chandler, 1984)

Ways of Coping Scale (Folkman & Lazarus, 1985)

Family Functioning (44)

Adolescent Family Life Satisfaction Index (Henry, Ostrander, & Lovelace, 1992)

Caregiver Strain Questionnaire (Brannan, Heflinger, & Bickman, 1997)

Child and Adolescent Burden Assessment (Messer, Angold, Costello, & Burns, 1996)

Children's Report of Parental Behavior Inventory (Schaefer, 1965)

Cleminshaw-Guidubaldi Parent-Satisfaction Scale (Guidubaldi & Cleminshaw, 1985)

Clinical Rating Scale for the Circumplex Model of Marital and Family Systems (Olson & Killorin, 1985)

Dyadic Parent-Child Interaction Coding system (Eyberg, 1974)

Family Adaptability and Cohesion Scales I (Olson, Bell & Portner, 1978)

Family Adaptability and Cohesion Scales II (Olson, Bell, & Portner, 1982)

Family Adaptability and Cohesion Scales III (Olson, Portner, & Lavee, 1985)

Family APGAR (Smilkstein, 1978)

Family Assessment Measure III (Skinner, Steinhauer, & Santa-Barbara, 1983)

Family Environment Scale (Moos & Moos, 1981)

Family Expressiveness Questionnaire (Greenberg, Mason, Lengua, & Conduct Problems Prevention Research Group, 1995)

Family Functioning Index (Pless & Satterwhite, 1973)

Family Functioning Questionnaire (Linder-Pelz, Levy, Tamir, Spenser, & Epstein, 1984)

Family Functioning Scales (Bloom, 1985)

Family Functioning Style Scale (Dunst, Trivette, & Deal, 1988)

Family Intervention Scale (Taynor, Nelson, & Daugherty, 1990)

Family Invulnerability Test (McCubbin, Olson, Lavee, & Patterson, 1985)

Family Life Questionnaire (Lautrey, 1980)

Family Relationship Inventory (Michaelson & Bascom, 1978)

Family Relationships Inventory (Bachman, 1970)

Family Satisfaction (Olson & Wilson, 1985)

Family Stressor Scale (Kessler, 1985)

Global Family Environment Scale (Rey et al., 1997)

Issues Checklist (Robins & Weiss, 1980)

McMaster Family Assessment Device (Epstein, Baldwin, & Bishop, 1983)

Modified Impact on Family Scale (Stein & Riessman, 1980)

O'Leary-Porter Scale (Porter & O'Leary, 1980)

Parent Perception Inventory (Hazzard, Christensen, Margolin, 1983)

Parent-Adolescent Attachment Inventory (Armsden & Greenberg, 1987)

Parent-Adolescent Communication Inventory or Scale (Barnes & Olson, 1982)

Parent-Adolescent Relationships Scale (Bell & Avery, 1987)

Parental Acceptance-Rejection Questionnaire (Rohner, Saavedra, & Granum, 1978)

Parental Nurturance Scale (Buri, 1989)

Parent-Child Relationship Inventory (Gerard, 1994)

Parent-Child Scales (Hetherington & Clingempeel, 1992)

Parenting Sress Index (Abidin, 1997)

Perceptual Indicators of Family Quality of Life Scale (Rettig et al., 1989)

Quality of Parent-Child Relationships Scale (author unknown)

Self-Report Family Inventory (Beavers, Hampson, & Hulgus, 1985)

Structural Family Interaction Scale-Revised (Perosa, Hansen, & Perosa, 1981)

Youth-Parent Conflict Scale (Prinz, 1979)

Family Resources (7)

Family Needs Scale (Dunst, Trivette, & Deal, 1988)

Family Needs Survey (Bailey & Simeonsson, 1985)

Family Resource Scale (Dunst & Leet, 1987)

FISC Family Needs Survey (McGrew, Gilman, & Johnson, 1989)

Home Index (Gough, 1949)

Prioritizing Family Needs Scale (Finn & Vadasy, 1988)

Support Functions Scale (Dunst, Trivette, & Deal, 1988)

Functioning Competence (10)

Adaptive Behavior Inventory for Children (Mercer & Lewis, 1982)
Assessment of Interpersonal Relations (Bracken & Howell, 1991)
Behavioral and Emotional Rating Scale (Epstein & Sharma, 1998)
CBCL and YSR Social Competence Scales (Achenbach, 1991a; Achenbach, 1991b)
Child Behavior Rating Scale (Rochester Social Problem Solving Core Group, 1980)
Health Resources Inventory (Gesten, 1976)
Instrumental and Social Competence Scale (Beiser, Lancee, Gotowiec, Sack, & Redshirt, 1993)
Normative Adaptive Behavior Checklist (Adams, 1984)
Responsibility and Independence Scale for Adolescents (Salvia, Neisworth & Schmidt, 1990)
Vineland Adaptive Behavior Scales (Sparrow, Balla & Cichetti, 1984)

Functioning Impairment (6)

Child and Adolescent Functional Assessment Scale (Hodges, Kline, Stern, Cytryn, & McKnew, 1982)
Children's Global Assessment Scale (Shaffer et al., 1983)
Columbia Impairment Scale (Bird et al., 1993)
Hampstead Child Adaptation Measure (Target & Fonagy, 1996)
Personal Problems Checklist-Adolescent (Schinka, 1985) and Children's Problems Checklist (Schinka, 1985)
Teacher Observation of Classroom Adaptation - Revised (Werthamer-Larsson, Kellam & Wheeler, 1991)

General Symptoms (19)

Achenbach scales	Child Behavior Checklist 4/18 (1991a)
	Youth Self-Report (1991b)
	Teacher Report Form (1991c)

ACQ Behavior Checklist (Achenbach, Conners, & Quay, 1983)
Behavior Dimensions Rating Scale (Bullock & Wilson, 1989)
Behavioral Problems Index (Peterson & Zill, 1986)
Behavior Rating Profile-2nd edition (Brown, & Hammill, 1990);
Bristol Social Adjustment Guides (Stott, 1969)
Burks' Behavior Rating Scales (Burks, 1977)
Devereux Scales of Mental Disorders (Naglieri, LeBuffe, & Pfeiffer, 1994)
Eyberg Child Behavior Inventory (Eyberg & Ross, 1978)
Louisville Behavioral Checklist (Miller & Roid, 1988)
Ontario Child Health Study scales (Boyle et al., 1993)
Parent Daily Report (Chamberlain, 1980)

Pediatric Symptom Checklist (Murphy & Jellinek, 1985)

Revised Behavior Problem Checklist (Quay, 1983)

Rutter's Parent Questionnaire (Rutter, Tizard, & Whitmore, 1970)

Vermont System for Tracking Client Progress (Burchard & Bruns, 1993)

Walker Problem Behavior Identification Checklist (Walker, 1976)

Goals (2)

Adolescent Goal Attainment Scale (Maher & Barbrack, 1984)

Family Goal Recording (Fleuridas, Rosenthal, Leigh, & Leigh, 1990)

Maltreatment (10)

Adolescent Abuse Inventory (Sebes, 1983)

Checklist for Child Abuse Evaluation (Petty, 1990)

Child Abuse Potential Inventory (Milner, Gold, & Wimberly, 1986)

Childhood Trauma Questionnaire (Bernstein et al., 1994)

Children's Impact of Traumatic Events Scale - R (Wolfe et al., 1991)

Child Report of Treatment Issue Resolution (Nelson-Gardell, 1997)

Child Sexual Behavior Inventory (Friedrich, 1990)

History of Victimization Form (Gentile, 1988)

Sexual Abuse Fear Evaluation Subscale (Wolfe & Wolfe, 1986)

Trauma Symptom Checklist (Briere & Runtz, 1989)

Multidimensional (17)

Adolescent Treatment Outcomes Module (Robbins et al., 1997)

Behavior Assessment System for Children (Reynolds & Kamphaus, 1992)

Child Behavior Rating Scale (Kilpatrick, 1975)

The Health of the Nation Outcome Scales for Children and Adolescents - Version 6.0 (Gowers, Whitton & Harrington, 1998)

Life Stressors and Social Resources Inventory-Youth (Moos, Fenn, Billings, & Moos, 1989)

Matson Evaluation of Social Skills with Youngsters (Matson, Rotatori, & Helsel, 1983)

Mental Status Checklist (Dougherty & Schinka, 1988; 1989)

Modified Rutter A2 Parent Questionnairre (Goodman, 1994)

Ohio Youth Problems, Functioning, and Satisfaction Scales (Ogles, Davis & Lunnen, 1998)

Paddington Complexity Scale (Hunter, Higginson, & Garralda, 1996)

Questionnaire developed for specific study (Slee & Rigby, 1993)

Scales of Independent Behavior (Bruininks, Woodcock, Weatherman, & Hill, 1984)

Service Utilization and Risk Factors interview (Goodman et al., 1998)

Social Adjustment Inventory for Children and Adolescents (Gammon, John, Prusoff, & Weissman, 1982)

Social Skills Rating System (Gresham & Elliot, 1990)

Strengths and Difficulties Questionnaire (Goodman, 1997)

Timberlawn Child Functioning Scale (Dimperio, Blotcky, Gossett & Doyle, 1986)

Quality of Life (7)

Child Health and Illness Questionnaire (Starfield et al., 1993)

Child Well-Being Scales (Magura & Moses, 1986)

Hopelessness Scale for Children (Kazdin & Petti, 1982)

Perceived Life Satisfaction Scale (Smith, Adelman, Nelson, Taylor, & Phares, 1987)

Quality of Life Profile: Adolescent version (Raphael, Rukholm, Brown, Hill-Bailey, & Donato, 1996)

Students' Life Satisfaction Scale (Huebner, 1991)

Quality of School Life Questionnaire (Epstein & McPartland, 1978)

Safety (1)

Feelings of Safety items (Schwab-Stone et al., 1995)

Satisfaction (4)

Client Satisfaction Questionnaire-8 (Larsen, Attkisson, Hargreaves, & Nguyen, 1979)

Consumer Satisfaction Questionnaire (author unknown)

Satisfaction Scales (Brannan, Sonnichsen, & Heflinger, 1996)

Youth Satisfaction Questionnaire (Stuentzner-Gibson, Koren, & DeChillo, 1994)

Self-Esteem (22)

Canadian Self-Esteem Inventory for Children (Battle, 1976)

Coopersmith Self-Esteem Inventories (Coopersmith, 1967; 1982)

Culture-Free Self-Esteem Inventories II (Battle, 1992)

Global Self-Worth scale (subscale of the Perceived Self-Competence Scale for Children, Harter, 1982)

Lipsitt Self-Concept Scale (Lipsitt, 1958)

Multidimensional Self Concept Scale (Bracken & Howell, 1991)

Offer Self-Image Questionnaire for Adolescents -Revised (Offer & Ostrov, & Howard, 1982)

Perceived Competence Scale for Children (Harter, 1982)

Perceived Self-Worth Questionnaire (author unknown)

Personal Attribute Inventory for Children (PAIC; Parish & Taylor, 1978); revised by Parish and Rankin (1982) and renamed the Nonsexist Personal Attribute Inventory for Children (NPAIC)

Piers-Harris Self-Concept Scale (Piers, 1969; revised manual, 1984). Also called: Children's Self-Concept Scale and The Way I Feel About Myself

Rosenberg Self-Esteem Scale (Rosenberg, 1965)

Self-Appraisal Inventory (Narikawa & Frith, 1972)

Self-Description Inventory (Mboya, 1993)

Self-Description Questionnaire I (Marsh, Parker, & Smith, 1983)

Self-Description Questionnaire II (Marsh, Smith, & Barnes, 1983)

Self-Description Questionnaire III (Marsh et al., 1990)

Self-Esteem Index (Brown, & Alexander, 1991)

Self-Perceived Role Competence (Beiser, Lancee, Gotowiec, Sack, & Redshirt, 1993)

Self-Perception Profile for Adolescents (Harter, 1988)

Self-Perception Profile for Children (SPPC; Harter, 1985) (Revision and extension of the Perceived Competence Scale for Children)

Tennessee Self-Concept Scale (Roid & Fitts, 1988)

What I Think About Myself (Mathewson, Orton, 1963)

Social Support (7)

Family, Friends, and Self Assessment Scales (Simpson & McBride, 1992)

Perceived Social Support Scale from Family (PSS-FA) and Perceived Social Support Scale from Friends (PSS-FR; Procidano & Heller, 1983); Perceived Social Support from School Personnel (DuBois, Felner, Brand, Adan, & Evans, 1992)

Personal History Inventory for Children (Parish & Wigle, 1985)

Social Support Appraisals Scale - Revised (Dubow & Ullman, 1991)

Social Support Scale for Children and Adolescents (Harter, 1985)

Supportive Parenting Scale (Simons, Lorenz, Conger, & Wu, 1992)

Stressful Events (9)

Academic Pressure Scale for Adolescents (West & Wood, 1970)

Adolescent Perceived Events Scale (Compas, Davis, Forsythe, & Wagner, 1987)

Children's Own Perceptions and Experiences of Stressors (Colton, 1989)

Daily Hassles Questionnaire (Rowlison & Felner, 1988)

Family Inventory of Life Events (McCubbin & Patterson, 1983)

Hassles Scale for Children (Parfenoff & Jose, 1989)

Life Events Checklist (Johnson & McCutcheon, 1980)

Life Events Scale for Children (Coddington, 1972)

Sandler and Block's (1979) Modification of the Coddington Life Events Scale for Children

Therapeutic Process (12)

Adolescent Working Alliance Inventory (Linscott, DiGiuseppe, & Jilton, 1993)

Barriers-to-Treatment Participation Scale (Kazdin, Holland, Crowley, & Breton, 1997)

Family Involvement items (Baker, Blacher, & Pfeiffer, 1993)

Loyola Child Psychotherapy Process Scales (Estrada & Russell, 1994)

Motivation to Change Questionnaire (Phares & Danforth, 1994)

Penn Helping Alliance Questionnaire (Alexander & Luborsky, 1986)

Psychotherapy Process Inventory (modified for use with children) (Gorin, 1993)

Smith-Acuna, Durlak, & Kaspar (1991) unnamed scales

Therapeutic Alliance Scales for Children (Shirk & Saiz, 1992)

Therapy Process Checklist (Weersing, 1996)

Treatment Acceptability Questionnaire (Hunsley, 1992)

Treatment Evaluation Inventory (Kazdin, 1980) and TEI-Short Form (Kelley, Heffer, Gresham, & Elliott, 1989)

Coping Skills

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Adolescent-Coping Orientation for Problem Experiences Scale (A-COPE; Patterson & McCubbin, 1987)	Measures how person relates with family members and peers to cope with stress; 54 items; 5-pt. Scale	Adolescent self-report	Development sample was middle to upper SES; results significantly affected by gender and ethnicity		Alpha of factors = .45 to .80	13 factors > 1.0 eigen values; loaded .362 to .836; accounted for 60% of variance; original study found 12 factors also explaining 60% of variance Corr. w/ Kidcope expected items = .11 to .62				
Adolescent Coping Scale (Frydenberg & Lewis, 1993); long and short forms	Assesses coping strategies, focusing on adaptive strategies and well-being			Long form (10 mins); short form (2-3 mins); manual from ACER (Australia)						Ages 12-18
Analysis of Coping Style (Boyd & Johnson, 1981)	Measures coping style child most likely uses in stressful situations; 20 pictures in peer and adult categories (6 statement choices per picture)			Semi-projective measure		Discriminated children w/ identifiable behavior probs and conduct disorders from others; 87.5-93.3% agreement of judges familiar with the coping styles model on statements matching coping styles				
Coping Inventory for Adolescents (Fanshawe & Burnett, 1991)	Assesses adolescent coping mechanisms				Alpha = .70; subscales = .67 to .77	Developed from A-COPE - this measure was factor analyzed - criteria - .4 factor loading, minimum 5 items per factor, eigen values >1; explained 51% of variance				
Coping Responses Inventory (Moos, 1997)	Identifies and monitors coping strategies (approach and avoidance); asks for actual and preferred coping	Self; individ or grp		10-15 min.; written at 6th grade reading level; Requires clinician with some training in administration and scoring of psychological tests; Intro kit = $67; $28/25 forms						Adult and youth versions (ages 12-18)
Coping Strategies Inventory (CSI; Tobin, Holroyd, & Reynolds, 1984)	Measures problem solving, cognitive restructuring, emotional expression, problem avoidance, social support, wishful thinking, social withdrawal, and self-blame 72 items (5 pt. scale)				2 week test-retest = .67 to .83. Alpha = .71 to .94 across scales.	8 factor solution; differentiates depressed vs. non-depressed; neurotic vs. non-neurotic; headache vs. non-headache In a study of the Kidcope, CSI subscales expected to correlate ranged from .33 to .77. Factors have been replicated in factor analysis. Content validity: good conceptual structure - developed from the WOC (Cook & Heppner, 1997).				Often used with college students

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Kidcope (Spirito, Stark, & Williams, 1988)	Measures coping skills; 10 items		Designed w/ middle class white persons; Has German version		3 day test retest for frequency = .56 to .75 for items; for efficacy = .25 to .74; 1 week frequency = .07 to .83; efficacy = .01 to .50; 2 week frequency = .04 to .56; efficacy = .21 to .58; 10 week frequency = .15 to .43; efficacy = .07 to .34	Content validity: items from literature on factor analysis of other coping measures; scale to 9th graders for scale development; factor analysis; then looked internal consistency - looked at results and threw the scale out and started over with 10 most common categories from the literature. Corr. of items w/ CSI expected factors = .33 to .77 Corr. w/ ACOPE expected items = .11 to .62 Significantly differentiated referred vs. controls and diabetic children vs. controls				
Problem Inventory for Adolescent Girls (Gaffney, 1984)	To assess social competence (response to situations); 52 items (5 pt. scale;					51 of 52 items differentiated delinquent and non-delinquent girls (p<.05); correlated .52 with IQ; correlated .32 with SES; correlated -.83 with a delinquent behavior checklist				Adolescent girls
Social Coping Questionnaire for Gifted Students (Swiatek, 1995)	Measures coping; 35 items (7 pt. scale)				Alpha = .54 to .79 across subscales					Gifted students
Stress Response Scale (Chandler, 1984)	Assesses child reaction to stress; impulsive acting out, passive-aggressive, impulsive overactive, repressed, dependent; 40 items (6 pt. scale); only measures maladaptive responses, no positive coping	Teacher or parent		10-15 min.; Requires clinician with some training in administration and scoring of psychological tests; Intro kit = $71; $32/25 answer and scoring sheets	Inpatient, outpatient, schools	1 month test-retest reliability = .86; .78 to .90 for subscales Alpha = .94	947 elementary school kids from 6 school districts in Western Penn.; children in special ed. Classes excluded	Experimental norms extend to age 18		Ages 5-14 in clinics, schools, & community agencies
Ways of Coping Scale (WCS; Folkman & Lazarus, 1985)	Measures one's appraisal of and definition of a self-selected stressful encounter; 8 subscales; 66 items; 4 pt. scale			Quite long for children (Spirito, Stark, & Williams, 1988)	Alpha = .59-.88 across subscales			Some differential use of coping strategies between younger and older adolescents		Ages 13-20

Family Functioning

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Adolescent Family Life Satisfaction Index (Henry, Ostrander, & Lovelace, 1992)	Assesses extent to which adolescents perceive family life in a positive manner; 13 items (5 pt. scale)		Tested in Southwest; Tested with 83% whites in one study; 6% native americans; 5% African American		Alpha = .85 to .90 across subscales Alpha for entire scale = .84 to .90	Content validity - 3 social scientists external from project evaluated whether items measured theorized construct (no data reported on this); minor revisions from use w/ 23 high school students Correlation between subscales = .36 Factor analysis revealed parent (Alpha = .88) and sibling factors (Alpha = .89) - both w/ eigen value >2; explained 62% variance; all items loaded on factor >.63 In second study, all items factor loaded on one of 2 factors >.50; factors explained 57% of variance Corr w/ family satisfaction scale = .67-.72 ; Parental scale corr w/ FSS = .70-.78 but Sibling subscale = .36-.43 across studies				Adolescents
Caregiver Strain Questionnaire (formerly Burden of Care Questionnaire; Brannan, Heflinger, & Bickman, 1997)	Assesses parental perception of stress experienced in caring for a child with serious emotional disturbance over past 6 months - negative feelings toward child, family's experience of burden, sense of worry, sadness, and fatigue; 21 items	Parent	Sample from military population - 72% white; 85% parent respondents female	10 mins	Cronbach's Alpha = .93 overall and .74-.92 for subscales	Cross & McDonald - exploratory and confirmatory factor analysis supported three factors - fit index = .96 correlated -.13 to -.30 with FAD correlated .20 to .42 with BSI except with Phobic anxiety scale; Cross & McDonald - needs more research			Cross & McDonald - score improves with child treatment	5-17 years old

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Child and Adolescent Burden Assessment (Messer, Angold, Costello, & Burns, 1996)	Measures family impact of emotional and behavioral problems; 20 items (4 pt. scale)	Interview w/ child's caregiver by lay interviewer		Training < 1 hour; Recc. to be administered after a diagnostic interview; 10 min.	Alpha >.84; 1 week test-retest ICC = .67; 12 mo. Test-retest = .32	Content validity - items selected after review of child and adult measures, and staff discussions; Factor analysis - 7 factors >1.0; examined scree plot and chose 1 factor solution; all items w/ factor loadings > .35 except substance use items; 27-30 % variance explained; Pearson correlation of similarity of factor loadings across 2 samples = .74 Clinical sample significantly higher scores than general sample; effect size of 2.2				
Children's Report of Parental Behavior Inventory (Schaefer, 1965)	Parental behavior toward child: acceptance, rejection, inconsistent discipline, control, hostile control; 42 items (3 pt. scale)	Child, parents	Tested with Hispanic children and mothers, hispanic mothers scored signif. diff on 4 of 5 scales vs. Anglo Americans for child and mother report		Alpha = .50 to .84	Corrs with other vars = -.58 to .50 Signif Discrims delinquents (half of the subscales), child guidance patients, an d maladjusted normals from controls Item selection guided by factor analyses of psychologists' ratings of parental behavior and review of conceptual models				
Cleminshaw-Guidubaldi Parent-Satisfaction Scale (Guidubaldi, Cleminshaw, 1985)	To identify troubled parent-child relationships and parent attitudes toward parenting; 5 domains: spouse support, parent-child relationship, parent performance, family discipline, & general satisfaction; 45 items	parent	Not related to employment status, income, religion	30 min.; requires masters degree to purchase; kit = $90; $29/25 forms	Internal reliabilities = .76 to .93	Factor analyzed to 5 factors; Strongly related to Dyadic Adjustment scale and Life Satisfaction Index	Has non-normalized standardized T scores and percentiles	Not related to age		Parents with any dependent child
Clinical Rating Scale for the Circumplex Model of Marital and Family Systems (CRS; Olson & Killorin, 1985)	Assesses structure of family coalitions (cohesion and adaptability subscales)	Clinician - semistructured family interview		Time varies; avail from Family Social Science: USA	Interrater corr = .88 for cohesion, .84 for adaptability, and .92 for communication	No validity studies as of 1987				All ages
Dyadic Parent-Child Interaction Coding System (Eyberg, 1974)	Measures parent-child interaction; 31 items			2 days	Interrater = .91 for parent behaviors; .92 for child behaviors	Significant differences btwn normal and problem children				

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Family Adaptability and Cohesion Scales I (FACES-I) (Olson, Bell & Portner, 1978)	Measures how family members perceive their families; 111 items; adaptability and cohesion dimensions, has social desirability scale			Original version that has been revised twice	Alpha = .75 for adaptability; .83 for cohesion	Factor analytic validity replicated original scales but unpublished Content validity - expert judges assigned items to dimensions Clinical relevance of items rated by 35 marriage and family counselors, then items selected based on ratings and loadings on factors Adaptability corr. w/ social desirability = .03 but cohesion corr. w/ sd = .45 58% normal fams rated as normal, 42% clinical fams rated as normal, 12-14% normal fams rated as chaotic/ disengated, 30-52% clinical fams rated as chaotic/ disengaged	Families with adolescent children = 201			Could not be used with younger children
Family Adaptability and Cohesion Scales II (FACES-II; Olson, Bell, & Portner, 1982)	Measures family dynamics; 30 items (5 pt. scale); factors: social desirability; togetherness; rules; autonomy	Fam members	Tested with Hispanic children - Cohesion score signif higher for Hispanic fams vs. Anglos	Shorter than FACES I, easier to read (especially for younger children)	KR-20 reliabilities = .48 to .91; Alpha = .72 to .87; 4 to 5 week test-retest = .83 for cohesion and .80 for adaptability Interspouse corr = .41-.46 cohesion, .25-.32 adaptability Parent/adolescent agree = .21 to .46	Discriminates extreme, mid-range, and balanced families in several problem categories; Content validity - expert judges assigned items to dimensions Signif discriminated adolescent repeat offenders and young adult prisoners from adolescent nonoffenders Corr. >.35 for 11 of 14 scales on FAD (not as predicted according to Halvorsen, 1991); Cohesion items load .35-.61; adaptability = .10-.55 Corr. w/ SFI = -.18 to -.78 for cohesion & -.39 to .02 for adaptability	Random national survey = 2082	Separate parent and adolescent norms		10 and up

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Family Adaptability and Cohesion Scales III (FACES-III; Olson, Portner, & Lavee, 1985); individual and couple versions	Measures family adaptability and cohesion; 20 items responded to twice - perceived and ideal family; 5 pt. scale Designed to reduce corr. btwn 2 scales which exist in FACES II	Family members, therapist	Hebrew and Norwegian versions, used in UK and Australia; Items do not reflect view of Hispanic, Black, or underpriveleged Anglo families (Bagarozzi, 1986)	Easy to score (Olson, 1986) 1-2 minutes to score; takes up to 20 minutes to complete	Alpha = .68; .77-.89 for cohesion and .62 - .72 for adaptability 4-5 week test-retest = .80 to .83 Cross - informant = 2x as many children see fams as disengaged compared to parent reports and published norms Corr btwn husband, wife, adolescent = .25 to .41 across subscales	Good face validity; content validity - scale based on circumplex model - developed from extensive literature review; items from factor analysis of FACES II Convergent: FACES III (COH scale) corr w/ FES (COH) = .86 SFIS-R (EN/D scale) corr w/ FACES III (COH) = .85 FAD (ROLES scale) corr w/ FACES III (ADAP) = .07 (bad) SFIS-R (FLEX scale) corr w/ FACES III (ADAP) = .40 Corr. w/ social desirability = .39 for cohesion, 0 for adaptability Discrim valid: Corr btwen Cohesion and adaptability = .03-.33 several studies Corr. for opp dimension across instruments: FACES III(COH) w/ SFIS-R(ADAP)=-.56 & .56 Corr w/ opp dim SFIS-R (COH) w/ FACES III(ADAP) = .15 to .28; Opp dim. FACES III w/ FES =-.23 to -.28 FAD (Probsol scale) corr w/ FACES III (COH) = .77(bad) differerentiated families with neurotics and schizophrenics from no-therapy groups; fams w/ LD child vs. fams without LD child; fams w/ child w/ learning and behavior probs from controls	2453 adults and 412 adolescents; norms for families with different age children, and in different stages of life cycle, cutting scores for each dimension			

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Family APGAR (Smilkstein, 1978)	Assesses functional state of patient's family: Adaptation, partnership, growth, affection, resolve; 5 or 20 items; 3-pt. scale		Has been studied in Taiwan, studied with non-representative samples; signif diff in scores depending on parental education, miniority status		Alpha = .80 - .86; Split half = .93 2 week test-retest = .83	Corr. w/ FFI = .80 Scores of clinical and normal groups signif diff. Identifying patients with psychological distress: Sensitivity = 68%, Specificity = 62%, false pos rate = 16%, false neg. rate = 19% Identifying PSC psychosocial probs - Kappa = .24; sensitivity = .30; specificity = .93; physician report psychosocial probs kappa = .14; sensit=.18; spec = .92; CBCL kappa = .20 Corr. w/ therapist rating = .64 Items selected from common themes in literature on families (Smilkstein, 1978)	None, has cutoffs with no empirical basis			Has been used 10 and up
Family Assessment Measure III (Skinner, Steinhauer, & Santa-Barbara, 1983)	Assesses family function, task accomplishment, role performance, affective expression, , affective involvement, control, values and norms; 134 items			30 -45 min.	Alpha = .27 to .87 for adult and child subscales	Corr. .44 w/ social desirabilty 4 discriminant functions w/ FAM signif discrim clinical and normal families Content validation - construct defined, items generated then rated for clarity, construct saturation, and clinical relevance	Normative sample = 247 adults and 65 adolescents from normal fams, over 2000 individs from clin fams			

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Family Environment Scale (Moos & Moos, 1981); child version (1984) available	Measures family social climate and level of agreement among family members; includes relationships, personal growth, and system maintenance; measures real, ideal, and expected family environment; 90 items; three forms (Real Form, Ideal Form, and Expectations Form; true-false format.	Any family member	Cross & McDonald - reliable with samples of African Americans and Hispanics but very low with Puerto Ricans or Vietnamese refugee youth - may not translate well, dissimilar value orientation, colloquialisms, negatively worded questions; Instrument designed using samples of blacks, Mexican-Americans	Time varies, approx 10 mins	Alpha = .10 to .89 across subscales and studies Kuder Richardson = .73. Mean item-item corr. = .20. Test retest at 4 months .54 to .91 for subscales; At 2 months = .68 to .86; At 1 year = .52 to .89 At 3 years = .47 to .58 At 4 years = .45 to .54 Hammil, Brown, & Bryant report <.80 for child version	Cross & McDonald: One study did not support subscales by confirmatory factor analysis Factor analysis found 2 factors only explain 28-34.1% of variance Studies find it varies in predicted direction according to family composition and presence of specific problems in family; Shares <.15% variance w/ MMPI family scales Corr w/ SFI = .17 to .73 Corr btwn SFIS-R (EN/D scale) w/ FES (COH scale) = .89 SFIS-R (FLEX scale) corr w/ FES (CONT) = -.59 FAD (Probsol scale) corr w/ FES (COH) = .78 FAD (AFFIN scale) corr w/ FES (COH) = .68 FACES III (COH scale) corr w/ FES (COH) = .86 correlated with staff member ratings; not found corr. w/ family card sort; FES cohesion w/ no corr. w/ other cohesion measures in a study Subscales corr. w/ social desirability = .02 to .44 Discrim valid - FES (COH) w/ FES (CONT) = .28 Corr w/ opp dim SFIS-R w/ FES= -.23 to -.65 Opp dim. FACESIII w/ FES =-.23 to -.28 most studies find that it differentiates treated from control families; Content validity -items from observations and interviews with families; items chosen for three conceptual areas that were defined first. 67% of graduate student raters placed 24 of 45 items on correct subscale; 67% of raters another sample placed 39 of 45 items correctly (this does not appear to be strong evidence?) Items selected based on psychometric properties of 200 item version given to large sample (item-response frequency, highest corr. w/ intended scale, equal # of true and false items per scale, low to mod scale intercorrs., max discrimination btwn fams) but results not provided (Forman & Hagan, 1983) Discriminated between high-service-use families and low-service-use families on 5 of 10 subscales.	Standard scores for normal and distressed families; child norms collected between 1976-1985 Sample 1468 adults, 621 children 1125 normal fams, 500 distressed fams	Authors say can use for any age group	Cross & McDonald - Conflict and family cohesion subscales found to respond to treatment	Child version from 5-12.11; Adult version - 10 and up
Family Expressiveness Questionnaire (Greenberg, Mason, & Lengua, & Conduct Problems Prevention Research Group, 1995)	Assesses family communication of positive and negative emotional states; 13 items	parent			Alpha = .82	2 factors that account for 52% of variance				
Family Functioning Index (Pless & Satterwhite, 1973)	To indicate whether families need assistance coping with stress, assesses marital satisfaction, frequency of disagreement; 19 items	parents		Unclear scoring criteria; 15 min.	Interspouse corr = .65-.72	Corr. w/ FAPGAR = .80 Mean score signif higher for clinical vs. normal fams Therapist-husband corr = .48; therapist-wife = .35-.39 Content validity - created from semistructured interviews with parents of school aged children and literature - not clear how items generated	No norms		5 year test-retest = .83	

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Family Functioning Questionnaire (Linder-Pelz, Levy, Tamir, Spenser, & Epstein, 1984)	Assesses family functioning; 8 items (5 pt. scale)				Alpha = .85 for both partners, .89 for only male from family; .28 when only female from family Interspouse corr. = .75	Items selected from lit review and operationalizing their definition of fam. func., discarded items not eliciting variation in response Face valid, no validation studies yet (Halvorsen, 1991)				
Family Functioning Scales (Bloom , 1985)	Assesses relationship and system maintenance factors of family functioning; 60 items				Alpha = .60 to .90 minus Organization scale	Corrs with other vars = -.31 to .25				
Family Functioning Style Scale (Dunst, Trivette, & Deal, 1988)	Evaluates family functioning; 26 items (4 pt. scale)	Checklist by family member or structured interview by clinician		92% of surveyed mothers felt it was appropriate in length but 56% of clinicians felt it was too long	Alpha = .92; split-half reliability = .92	Corr w/ Family Hardiness Scale = .74 Factor analysis = 5 factors explaining 60% of variance; only 1 item did not load				
Family Intervention Scale (Taynor, Nelson, & Daugherty, 1990)	Measures fam func: role performance, connectedness, physical maintenance, community resources, maintaining emotional wellbeing; 5 items (11 pt. Likert scale)	Service providers	Tested on very small sample in Ohio	2 to 3 min.	All items to total score Pearson r's >.77 Interrater reliability, Pearson r = .78 for 11 families for total score; for subscales = .04 to .88; the low subscale was not dropped	Items generated from interviews with service providers, review of target areas with agency providers Signif discriminated fams rated by senior supervisory counselor as high or low functioning (1st study done so as to get agency interested in use of and research of measure)			Items showed differential response to intervention in manner expected by authors; scale found to detect signif change following tx	
Family Invulnerability Test (McCubbin, Olson, Lavee, & Patterson, 1985)	Measures family stresses, strengths, and adaptability to the environment						1000 family normative sample			
Family Life Questionnaire (Lautrey, 1980)	Assesses harmony in family life; 24 items				Alpha = .84 to .91; test-retest = .61 to .84	Corr w/ Mother-Daughter Relationship scale = .40				
Family Relationship Inventory (Michaelson & Bascom, 1978)	Clarifies individual feelings and interpersonal relationships; 50 items (half positive and half negative items)			30 min.; simple vocabulary; recc. for psychologists, sw's, family counselors, youth counselors; Requires clinician with some training in administration and scoring of psychological tests; Introductory kit = $89						5-adult
Family Relationships Inventory (Bachman, 1970)	Measures parent-child interaction, 26 items, rating scale format	Child			Alpha ranged from .66-.84 across subscales					

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Family Satisfaction (Olson & Wilson, 1985)	14 items (5 pt. scale)	Parent, adolescent			Cronbach's Alpha = .90-.93 across studies Spouse agree = .35 Parent-adolescent agree = .29 to .31	Corr w/ Adolescent Family Life Satisfaction Scale = .67-.72;				
Family Stressor Scale (Kessler, 1985)					Alpha = .66					
Global Family Environment Scale (GFES; Rey et al., 1997)	Measures the "family's ability to provide physical and emotional care, secure attachment relationships, consistency, and appropriate, nonpunitive limits" (Rey et al., 1997, p. 817), score ranges from 1-90 where 81-90 indicates adequate family environment, 71-80 slightly unsatisfactory, 51-70 moderately unsatisfactory, 31-50 poor, 11-30 very poor, and 1-10 very disturbed.	Clinician	Validated in 4 Australian mental health settings and 1 setting in Hong Kong, China; interrater reliabilities did not differ by nationality; correlates with socioeconomic status (.40)	At least minimal training (1 hour) required, although standard training (2-3 hours) resulted in small but significant improvement to interrater reliability (see Rey et al., 1997)	Interrater agreement (ICC) in a preliminary study using 20 case charts was .90 across 4 raters. Interrater agreement (ICC) with 7 case vignettes averaged .82 across 5 settings (range .74-.91) at Time 1, .84 (range .75-.96) at Time 2. Average interrater agreement across 4 settings with 211 clinical patient interviews was .89. Interrater agreement was significantly less with minimal training (.84) than with standard training (.93); test-retest average across 5 settings was .91.	Construct validity: t-tests revealed lower scores among children with diagnoses of oppositional defiant and conduct disorders; concurrent validity: correlation with the Psychosocial Adversity Index (PSAI; Biederman et al., 1995), an index of family environment risk factors for ADHD, = .76				
Issues Checklist (Robins & Weiss, 1980)	Measures conflict between parents and adolescents; 44 items (5pt. scales)				Alpha = .61 to .71 across subscales					

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
McMaster Family Assessment Device (FAD; Epstein, Baldwin, & Bishop, 1983)	To identify family system problems; general functioning and 6 subscales: problem-solving, communication, roles, affective responsiveness, affective involvement, and behavior control; 60 items, 4 pt. scale	Any family member older than 12	Samples primarily white, middle class, and from intact families; Large part of sample from students in Introductory Psychology course Research with Hawaiian-Americans and Japanese-Americans suggest need for modification; SES found to significantly influence results	15-20 mins, avail from Brown University Family Research Program, USA	Cronbach's Alpha =.86 to .92 and Gutman split half coefficient = .83 Cronbach's alpha for subscales from .70 to .86 .except for Roles scale .57 to .69 Alpha = .72 to .92; mean = .78 Test-retest (1 week) = .66-.76	Corr -.06 to -.19 with measures of social desirability; Factor analysis supports hypothesized structure across several samples, signif discrims families rated by clinicians as healthy and unhealthy; Identified correctly 67% non-clinical grp and 64% clinical group; Signif Discrims normal and clinical fams on 6 of 7 scales For health/pathology cutoff scores (not based on published norms): sensitivity = .57 to .81, specificity = .60 to .79 Cross & McDonald -Significantly associated with parental police involvement, alcohol abuse in family, parental emotional disorder, marital violence, parental separation, family structures Corr >.50 for 6 of 8 scales on Family Unit Inventory; >.35 for 11 of 14 scales on FACES -II (not as predicted according to Halvorsen, 1991); Corr. w/ SFI = .07 to .77 FAD (AFFIN scale) corr w/ SFIS-R (FCC/E scale) = .49 Affective Involvement scale correlated .65 to .68 with cohesion scales of FACES III and FES; Discrim valid - Scales do not discriminate btwn cohesion and adaptability. Corr cohesion and adaptability scales= .44 to .67 FAD adaptability scales correlated ..07 to .63 with adaptability scales (FACES III, SFIS-R, FES) but >.60 with family cohesion - FES, FACES III(bad); FAD (COH) corr. w/ SFIS (ADAP) = -.63, .59, w/ FES (ADAP) = -.35, w/ FACES III (ADAP) = .22 Sample sizes too small (Halvorsen, 1991)	Standardized w/ 503 members			Any family member over age 12
Modified Impact on Family Scale (Stein & Riessman, 1980)	Assesses effects of behaviorally difficult children on their families; 23 items (5 pt. scale)		Has italian version		Cronbach's Alpha = .93; split-half-reliability = .96; 4 day test-retest = .96	Correlations with other variables = .66 to .71				
O'Leary-Porter Scale (Porter & O'Leary, 1980)	Assesses interparental conflict in the presence of children; 10 items; 5 pt. scale		Tested with African American families		Internal consistency = .77 to .87					
Parent Perception Inventory (Hazzard, Christensen, Margolin, 1983)	Assesses warmth and support between family members; 18 items (5 pt. scale)				Alpha = .78 to .88	Corr w/ child's self-esteem, and parent report of child conduct disorder - no data				
Parent-Adolescent Attachment Inventory (Armsden & Greenberg, 1987)	Assesses adolescents' perception of quality of relationships to parents; 6 items (5 pt. scale)	Adolescent			Alpha = .81 to .86					

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Parent-Adolescent Communication Inventory or Scale (Barnes & Olson, 1982)	Assesses the qualities of open and problem parent-child communication; 20 items (5 pt. scale)	Parents and kids	Tested with Hispanics - No diff from Anglos on mother report, one scale signif lower for Hispanics on child report Korean version used		Alpha = .60 to .91 across studies and reporters and scales Spouse agree = .30 Parent adolescent agree = .32 to .34	Corr w/ Behavioral adjustment = -.22 to -.49				Designed for adolescents but authors suggest can be used with younger children
Parent-Adolescent Relationships Scale (Bell & Avery, 1987)	Measures closeness to parents, 21 items, 5-point scale	Adolescent			Alpha ranged from .71-.78 across 2 subscales					
Parental Acceptance-Rejection Questionnaire (Rohner, Saavedra, & Granum, 1978)	Measures subjective perceptions of acceptance-rejection by caregivers, 60 item (4 pt. scale)	Self-report - child and adult retrospective and parent views of their behavior toward child versions	"Common factor structure across 8 sociocultural groups worldwide" (Cournoyer & Rohner, 1996); In Bengali sample, several items did not load on the logical scale raising questions of translation flaws Used with Korean adolescents Translated into 17 languages		Alpha coefficients = .72 - .95 for scale scores, simple zero-order correlation between childhood and adolescent scores (7 years apart) was .62	Correlations with criterion measures (not specified in Cournoyer & Rohner, 1996) ranged from .43-.90 "Common factor structure across 8 sociocultural groups worldwide" (Cournoyer & Rohner, 1996) - explain 27 to 43% of variance (1st 2 factors) - beyond 2 factors any factor only explains 3% or less of variance (low variance explained) Discrim validity - across 8 studies, median corr btwn factors = -.38				
Parental Nurturance Scale (Buri, 1989)	Measures parental nurturance; 24 items; has mother and father versions	Child			Alpha = .92 to .95; Test-retest = .92 to .95	Corrs with self-esteem = .44 to .67				
Parent-Child Relationship Inventory (Gerard, 1994)	Assesses how parents view the task of parenting and how they feel about their children; 78 items; 2 validity scales	Parent self-report		4th grade rdg. level; 15 min.; kit=$95; $33/25 forms			Standardized on 1100 parents; separate norms for mothers and fathers			Parents of 3 - 15 y.o. children

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Parent-Child Scales (Hetherington & Clingempeel, 1992)	Measures parent-child relationships - 11 scales: parent-child warmth - 6 items (5 pt. scale), discipline - 24 items (5 pt. scale), Behavioral support - 20 items (7 pt. scale), and conflict - 32 items (7 pt. scale);	Mothers, fathers, children			Internal consistency on subscales = .62 to .96; 2 month test-retest = .63 to .87	Claim validity supported by multitrait-multimethod analysis using cross informant correlations and independent observations - expected subscales more closely related to one another; However, across informants, some subscales more highly correlated with other subscales than with the same dimension by a dift. informant; other subscales seem to have good convergent and discriminant validity - especially conflict Differentiates between parent-child relationships in married, divorced, and remarried families; Correlated with indices of children's psychological adjustment				
Parenting Stress Index (revised version) (PSI; Abidin, 1997	Measures caregiver strain; has child domain with 6 subscales (47 items): adaptability, acceptability, demandingness, mood, distractibility/hyperactivity, and reinforces parent; parent domain (54 items): 101 items plus 19-item Life Stress Scale (life events); short form has 36 items	Parents	Has Hispanic norms, has French Canadian version	20-30 mins; 10 min short form; requires 5th grade reading level; Requires clinician with some training in administration and scoring of psychological tests; minimum Master's degree; Introductory kit = $89.50; $34/25 forms; Software available from Multi-Health System, Inc. ($395)			Has 47 research based clinical profiles based on reference groups; norms by ages; normed on more than 2500 parents	Norms by age		Parents of children ages 12 and younger
Perceptual Indicators of Family Quality of Life Scale (Rettig, et al., 1989)	Measures respondents perceptions of the degree needs met within families; 35 items				Theta = .97; unknown type of reliability					
Quality of Parent-Child Relationships scale (author unknown)	Measures quality of mother-child and father-child relationships; 7 items (4 pt. scales)				Alpha = .86 mothers; .87 fathers					

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Self-Report Family Inventory (, Beavers, Hampson, & Hulgus, 1985)	Family members' perceptions of their family's functioning in 6 domains; suggested to use with Beavers-Timberlawn Family Evaluation Scale and the Centripetal/ Centrifugal Family Style Scale; 36 items, 5 pt. scale			Southwest Family Institute, USA, 1985	Alpha = .84 to .88; 30 to 90 day test-retest = .44 to .88 across studies	Corr. .09 to .61 w/ Locke-Wallace marital satisfaction scale; -.18 to -.78 w/ FACES II cohesion & -.39 to .02 for adaptability Corr. w/ FAD = .07 to .77 Corr w/ FES = .17 to .73 Corr w/ Bloom's family functioning scales = .26 to .86 Corr. w/ observational scores by therapist = .60 All items load >.50 on a factor Discrim valid - Cohesion scales to adaptability scales = .41 to .61 in MMMT analysis - higher corr. w/ cohesion scales than adaptability for some but not all scales Corr. for opp dimension across instruments: FACES III (COH) w/ SFIS-R (ADAP)=-.56 & .56 Corr w/ opp dim SFIS-R (COH) w/ FACES III (ADAP) = .15 to .28; Corr w/ opp dim SFIS-R w/ FES= -.23 to -.65 Unclear how items were selected	No norms			
Structural Family Interaction Scale-Revised (Perosa, Hansen & Perosa, 1981)	Identifies family interactions with 7 subscales: enmeshment-disengagement, flexibility-rigidity; family conflict avoidance-expression; mother-child cohesion-estrangement, father child cohesion-estrangement, spouse conflict resolved-unresolved, and cross-generational triads-parent coaltion; 68 items; 4-pt. scale				4 week test-retest = .81 to .92. Alpha = .71 to .94.	Corr btwn SFIS-R (EN/D scale) w/ FES (COH scale) = .89; SFIS-R (FLEX scale) corr w/ FES (CONT) = -.59; FAD (AFFIN scale) corr w/ SFIS-R (FCC/E scale) = .49; FAD (Probsol scale) corr w/ SFIS-R (FLEX scale) = .61; SFIS-R (EN/D scale) corr w/ FACES III (COH) = .85; SFIS-R (FLEX scale) corr w/ FACES III (ADAP) = .40 Discrim valid = SFIS-R (ADAP) w/ FACES III, FAD, & FES cohesion scales = -.45 to-.70 & .44 to .61 (bad) SFIS-R (COH) w/ FACES III & FES adaptability scales = .15 to .28 and -.23 to -.28 has differentiated family variables associated with clinical pops from those related to controls Discrim incest families from other clinical family controls or nonclinical families.				
Youth-Parent Conflict Scale (Prinz, 1979)	Measures youth-parent conflict; 20 items (true-false)	Parent and adolescent report			Alpha = .86					

Family Resources

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Family Needs Scale (Dunst, Trivette, & Deal, 1988)	Measures basic rescources, child care, personal family growth, financial and medical resources, child ed and therapy, food, etc.; 41 items (5 pt. scale)	Checklist by parent or by clinician		88% of surveyed mothers thought it was appropriate in length; but 52% of clinicians felt it was too long	Internal consistency: Alpha = .95; Split-half=.96	Content validity - broad representation of construct = 15 categories but only 2.7 items/category; Corr = .28 w/ parent internal locus of control; .42 w/ parent well-being Factor analysis = 9 component w/ eigen values >or = 1.0 (lenient criteria)	None			
Family Needs Survey (Bailey & Simeonsson, 1985)	Checklist indicating need for different types of resources such as financial or medical; 35 items (3 pt. scale)	Checklist by family member or structured interview by clinician		92% of surveyed mothers thought it was appropriate in length; but 48% of clinicians felt it was too long	6 mo. Test-retest = .52 to .81 Corr mother & father =.47 to .52 for subareas = .28 to .68	A study found 70% of targeted family goals from FNS Factor analysis found 5 factors; but five items did not load on any factor Only 60% of top 5 needs identified in open ended questions also named in scale Financial needs subscale signif discrim low from middle/upper income mothers	None			
Family Resource Scale (FRS; Dunst & Leet, 1987)	Measures parental perceptions of the adequacy of household resources (e.g., time, energy, money); 31 items (5 pt.scale)				Internal consistency: Alpha = .92; Split-half=.95 2-3 mo. Test-retest = .52	Content validity - covers 11 categories; less than 3 items/category; developed based on human ecology, social support, family systems theory and help seeking literature; Corr .81 btwn professional rank orderings and order of resources on scale (from basic up); .75 corr btwn professionals Corr = .45 to .57 on parent well being; .63 w/ parent commitment to intervention; .39 w/ SES; -.44 w/ mode of child birth, -.50 w/ birth related probs, .44 w/ informal social support, .53 w/ mother commitment to child care, .62 w/ frequency of social network contact 7 subscales account for 47% variance in parent wellbeing, and 48% variance in parent commitment to intervention Factor analysis - 8 components w/ eigen values > or = 1; all items factor loaded >.40 and accounted for 75% of variance	None			
FISC Family Needs Survey (McGrew, Gilman, & Johnson, 1989)	59 items			Cannot find additional data on this measure		Content validity - covered 13 categories of family need; over 4 items/category	None			
Home Index (Gough, 1949)	Measures socioeconomic status, 22 items, factual questions	Child				13 items validated against known father's occupation				

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Prioritizing Family Needs Scale (Finn & Vadasy, 1988)	Measures basic needs, health and safety, social needs, personal needs, needs for self-fulfillment; 41 items (7 pt. scale)			Can not find additional data on this measure		Content validity - examined correspondence btwn items and Maslow's heirarchy of needs; broad representation of construct = 17 categories but only 2.4 items/category	none			
Support Functions Scale (Dunst, Trivette, & Deal, 1988)	Measures emotional support, child support, financial support, instrumental support, agency support; 20 items (5 pt. scale)				Internal consistency: Alpha = .87; Split-half=.88 1mo. Test-retest = .62	Content validity - only covered 8 needs categories and only 2.5 items/category Corr = -.20 on time demands; .25 on family well being, .33 on personal well being Factor analysis - 5 components w/ eigen values > or = 1	none			

Functioning Competence

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Adaptive Behavior Inventory for Children (Mercer & Lewis, 1982)	Measures child's performance in social roles within the family, peer group, and community	Interview of parent or guardian	Spanish version	Kit = $107; $107/100 forms; 45 min.			Has standard scale scores			Ages 5-11
Assessment of Interpersonal Relations (Bracken, & Howell 1991)	Assesses quality of interpersonal relationships - family, social, and academic contexts			20 - 40 min.; $84/Intro kit; $42/50 forms	Internal consistency = .94 to .96 across subscales; .96 total scale; 2 week test-retest = .98		2,501 btwn 9-19 years old			
Behavioral and Emotional Rating Scale (BERS; Epstein & Sharma, 1998)	Measures behavioral and emotional strengths in the past 3 months with 5 subscales: interpersonal strengths, family involvement, intrapersonal strengths, school functioning, and affective strength; overall strength quotient; 8 open-ended items; 52 items; 4 pt. scale	An examiner administers the measure to parents, family members, teachers, or professionals who have knowledge of the child's behavior	Ethnicity norms (Caucasian, African American, and Hispanic)	10 mins; requires some training in psychological test administration and scoring; $74 for intro kit; $34/50 forms	Internal consistency averages for subscales across both normative groups = .84-.96, overall strength quotient = .97-.98; test-retest = .85-.99 across 5 subscales, = .99 on overall strength quotient; interrater reliability = .83-.96 across 5 subscales, = .98 on overall strength quotient; scoring agreement = .99 for each scale	Content validity: 250 professionals surveyed and generated 1,200 behaviors (190 non-redundant in 15 categories); 400 professionals rated relevance of 190 items to child strengths (63 lower scoring items were excluded); 37 of 127 items did not discriminate between children with and without SED, 10 were excluded due to infrequent membership in cells; factor analysis resulted in a 68-item, 5-factor scale; using an item discrimination validity cut-off = .3, 16 items were deleted. Item discrimination validity of the 52-items scale was computed with the two normative samples, ranging from .38-.80 across subscales and age. Concurrent validity: subscale correlations with Walker-McConnell scales = .29-.85; Self-Perception Profile for Children subscales = .28-.72; Teacher Report Form subscales = .28-.75. Construct validity: t-tests significantly distinguished between EBD and non-EBD on each scale (children with EBD scored lower on strengths); interrelationships between subscales = .40-.80 (EBD group), between subscales and strength quotient .57-.92; for NEBD group, .62-.87 and .75-.95	Nationally representative, standardized sample of 2,176 students ages 5-18.11; separate normative group of 861 children with emotional or behavior disorders; age, gender, and ethnicity norms			Ages 5-18.11 in schools, clinics, child welfare agencies

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
CBCL & YSR Social Competence Scales (Achenbach, 1991a; Achenbach 1991b)	Measures social competence over past 6 months and subscales - activities, social, school; 20 items; 14 for YSR;		Normative sample - 73% white, 16% black, 7% hispanic, 4% other, 81% of sample middle or upper class, no separate norms for ethnic grps or social classes because differences only explained small portion of variance (but has been suggested that it is not sensitive to functioning of impoverished or those in dysfunctional families); translated into 33 languages, has been studied in 15 different cultures; significantly larger portion of community children in clinical range for Puerto Rico sample	5-10 mins; Written at 5th grade reading level;	Internal consistency .42 to .64 for CBCL and YSR 1 week Test-retest intraclass correlation coefficient =.87 to .996 1-2 year test-retest = .56 to .63; unknown time period test-retest = .74 CBCL Inter-parent = mean of .79; Interrater reliability when used as an interview instrument = .93 YSR - test retest strong for older adolescents but less for younger; mean = .80 7 month test-retest of YSR = .62 TRF - good test-retest	Activities and school subscale not found to load on a social competence factor; Did not indicate adequate and inadequate social functioning based on a teacher report scale Corr w/ SAICA = .52 Discriminated clinic referred from controls (criterion related validity) - CBCL; YSR not as strong of a descriminator but significant 61% referred and 16% nonreferred above clinical cutoff TRF also distinguishes referral status Cross & McDonald criticize that it leaves out coping and altruism Construct validity of YSR not demonstrated - Salvia & Ysseldyke	Each sex, 4-11, 12-18; sample of 2,368 children; YSR norms 11-14 and 15-18	YSR - test retest strong for older adolescents but less for younger;	Cross & McDonald - has been used successfully to measure the effects of many interventions;	Children and adolescents in treatment
Child Behaviour Rating Scale (Rochester Social Problem Solving Core Group, 1980)	Measures children's psychosocial development; 28 items				Test-retest = .72 to .92					
Health Resources Inventory (Gesten, 1976)	School related competencies; 44 items (5 pt. Scale)	Teacher, parent			Test - retest = .72 to .91 across subscales; 10 mo. Test-retest = .71 to .78 Alpha across subscales = .79 to .88; Corr btwn teacher and mother = .32	5 factors accounting for 52 % of variance; Corrs among factors = .09 to .38 Corr. mother's ratings w/ CBCL dimensions = -.40 to -.48				

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Instrumental and Social Competence Scale (Beiser, Lancee, Gotowiec, Sack, & Redshirt, 1993)	Measures child's ability to get along with peers and perform appropriately at school; 25 items for parent; 21 for child	Parent, child	Tested with US and Canadian samples of Anglos, African Americans, American Indians, an Puerto Ricans - high internal consistency in these groups??	5 min.	No test-retest data Anglos, African Americans, American Indians, and Puerto Ricans - high internal consistency in these groups??	Adequate convergent validity?		Has signif corrs with something? across grade levels		
Normative Adaptive Behavior Checklist (Adams, 1984)	Measures self-help, home living, independent living, social skills, sensory motor, and language concepts		Hammil, Brown, & Bryant report unacceptable demographics of normative sample	22 min.	Internal consistency >.90 for total; one subscale below .80; test-retest <.80	Hammil, Brown, & Bryant report as not validated	Normed on at least 1000 subjects with 100 subjects per age; collected btwn 1976-1985			1 to 21-11
Responsibility and independence scale for adolescents (Salvia et al., 1990)	Measures adaptive behavior, responsibility, and independence		Hammil, Brown, & Bryant report acceptable demographics of normative sample	40 min.	Internal consistency >.80; test-retest >.80	Hammil, Brown, & Bryant report at least 5 types of empirical validation	Normed on at least 1000 subjects with at least 100 per age			12 to 19-11
Vineland Adaptive Behavior Scales (Sparrow, Balla & Cichetti, 1984)	Measures adaptive behavior across four domains (Communication, daily living skills, Socialization, motor skills) and subdomains; 3 versions: (a) Survey Form for screening, placement, and diagnosis; (b) Expanded Form for developing educational or treatment plans, and (c) Classroom Edition, Semi-structured interview; trained interviewer asks about the child's involvement in areas of adaptive behavior and probes for specific information; 577 items; 297 item version; and 45 item screener	Interview of parent or primary caregiver	Translations available in Spanish and several other languages; no diff in scores btwn European, Asian, and Afro-Carribbean children (N=70); some items are culture bound	Survey form: 30- 60 min.	Test-retest "excellent for subdomains, domains, and composite scores,"; Hammil, Brown, & Bryant report test-retest <.80 Agreement between mothers (Survey Form) and teachers high (not specified); Internal consistency >.80	In Sparrow & Cicchetti (1985): "studies investigating construct, concurrent, factorial, differential, and predictive validity...were also successfully conducted" Hammil, Brown, & Bryant report at least 3 types of empirical validation	National US norms; at least 1000 subjects with at least 100 subjects for each age; normed between 1976 and 1985	Has different items for different age groups		0-19 years, low-functioning children and adults

Functioning Impairment

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Child and Adolescent Functional Assessment Scale (Hodges, Kline, Stern, Cytryn, & McKnew, 1982)	Assesses child functioning impairment in the following areas: role performance, thinking, behavior toward others/self, moods/ emotions, and substance use; assesses caregiver on basic needs (ability to provide), family/social support; rating of most severe level in pre-specified time period (such as past month in FBEP); 4-point scales; has telephone interview and life functioning data sheet to aid in collection of additional information; service provider judgment may be subjective and biased, especially when the provider has a stake in the outcome	Service provider	"…no significant effects of race or caregiver education on CAFAS scores" (Hodges & Wong, 1996, p. 458), although another study (Hodges & Wong, 1997) reported slightly higher CAFAS scores for Caucasians and children from low income families	Approx 10 mins; 30 mins with telephone interview; 45-90 mins according to another source. As done in the FBEP, lay raters can be used; in the FBEP, training entailed 11/2 days to achieve satisfactory reliability. Hodges and Wong (1996) reported use of self-training manual; no training time specified	Test-retest = .95 (total score), .79 to .91 for subscales. Alpha = .63 to .68 across 4 administrations. Interrater reliability ICC coefficient using case vignettes was .40 in one study (Cross & McDonald, 1995). Interrater reliability (ICC) across 4 training samples w/ case vignettes: .63 to .92 (lay undergraduates), .65 to .89 and .71 to .88 (graduate samples), and .54 to .96 (agency staff); across domains: .84 to .89 (total), .68 to .79 (role performance), .74 to .79 (behavior), .63 to .75 (moods/ emotions), .87 to .96 (substance use), and .54 to .74 (family/social support); Cohen's Kappa for agreement on level of severity = .83.	Cross and McDonald (1995): Corr -.72 to -.81 with CGAS; discriminated between residential and outpatient grps; better predictor of restrictiveness of care, type and cost of services vs. PCAS and CBCL; needs more validity studies; "Corr moderately" with CBCL, BCQ, C-GAS, and CAPA; identified children with psychopathology according to PCAS; at intake, predicts service use at 6 and 12 months; higher impairment scores related to restrictive services, costs, bed days, and days of service. Hodges & Wong (1996): across 4 administrations, corrs = .42 to .49 (CBCL), .52 to .56 (CAS), .58 to .63 (PCAS), and .36 to .43 (BCQ).		Separate version for ages 4 to 7	Total score decreased over time in treatment	Grades 1-12

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Children's Global Assessment Scale (CGAS; Shaffer et al., 1983): Clinician and lay interviewer versions	Measures global functioning with one overall score; time frame specified by administrator; score from 1 to 100; behavior-oriented descriptions for every 10 points (lack of formal training materials); score of less than 70 suggested as a "case"	Clinician or lay interviewer-child or parent as interviewee	Reliability with sample 76% African American, 10% Hispanic; also tested with samples from several countries and of different religious backgrounds	Bird, et al. claim minimal training needed	ICC clinician test-retest = .91; .75 for lay interviewer Several studies with traditional mental health clinicians with interrater intraclass correlations .53-.93; Clinician corr w/ lay interviewer C-GAS Pearson = .68; parent C-GAS w/ lay interviewer C-GAS ICC= .65 One applied field study with poor interrater reliability; Cross and McDonald (1995) - satisfactory test-retest; Test-retest, interrater kappa on case vs. noncase = .65; reliability has yet to be established in applied settings; Interrater test-retest intraclass correlation = .83 Test-retest (6 mos) with identical case vignettes = .85. Agreement between attending psychiatrists and milieu staff (Pearson r) = .62 and .76, respectively; Agreement on caseness = 83.5% (kappa = .65); kappa = .64 in another study; test-retest ICC (average 19 days) with different clinicians = .72; ICC between observers (times 1 and 2) and video tape observers were .86 and .87, and .83 between video tape observers; Pearson rs with own Axis V ratings were -.92 and -.87, *M* = -.82 with other raters.	Significantly discriminated between cases and non-cases, cases and non-cases in the community sample, referred and non-referred, inpatient and outpatient (several findings); Regression models show that it significantly predicts service utilization -.62 to -.65 corrs with CBCL but .57 to .58 corrs with CBCL social competence; -.25 corr with Connors; -.80 corr with CAS; .25 corr. w/ Rutter Parent; .49 at 5 months post treatment Pearson corr w/ Axis V of DSM-III = -.82 to -.92 Pearson corrs w/ psychiatrist severity ratings = -.76 to -.90 85% agreement with discriminant function used to predict caseness Cut score of 61 with sensitivity to caseness of .85 and specificity = .80; Area under the Curve = .88; AUC for parental perceived need = .69 In a clinical practice study, corrs with CBCL total, internalizing, and externalizing scores ranged .00-.18. Corrs with 3 CBCL competence subscales ranged .13-.29. Corr .43 with WISC-R Full-Scale IQ, .33-.39 with Timberlawn Child Functioning Scale (social relatedness), and .00-.05 with family stress. Pearson rs with CBCL = -.65 (current) and -.62 (6 mo), with CBCL social competence = .58 and .57; discriminated between referred (*M* = 55) and non-referred (*M* = 78) and cases (*M* = 62) and non-cases (*M* = 86). Discriminant function correctly classified 85% of a sub-sample of cases. Discriminated between outpatients (*M* = 65) and inpatients (*M* = 46); corr .25 with Conners Abbreviated Parent Checklist (outpatients). Agreement with PSC caseness kappa = .82		Scores not found to be related to age	Cross and McDonald (1995): fairly strong sensitivity to change; Sourander, et. al, 1996: detected significant change over time Weissman et al. (1990) reported greater mean change (13 pts) for children with first onset of depression compared to children with no first onsets (4 pts)	4-16 years of age

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Columbia Impairment Scale (Bird, et al., 1993)	Assesses functional impairment - interpersonal relations, broad psychopathology domains, job or school functioning, use of leisure time; 13 items (5 pt. scale)	Parent or child		5 min.	Test-retest ICC = .89 for parent; .63 for child	Significantly discriminated children who had used mental health services from those who had not. Signif differentiated physical abused vs. not physically abused Corr. w/ clinician C-GAS = .63 for parent, .43 for child version.				
Hampstead Child Adaptation Measure (Target & Fonagy, 1996)	Assesses 14 dimensions of children's psychosocial adaptation	Parent								
Personal Problems Checklist-Adolescent (PPC-A; Schinka, 1985) and Children's Problems Checklist (CPC; Schinka, 1985)	PPC-A: Assesses problems in 13 areas: social, job, home, school, money, religion, emotions, family, appearance, dating, health, attitude, and crises; 240 items. CPC: Assesses problems in 11 areas: emotions, self-concept, peers/play, school, language/ thinking, concentration/ organization, activity level/motor control, behavior, values, habits, and health; 190 items	Adolescent; CPC completed by parent or guardian		PPC-A at 7th grad rdg level; does not require any training; $37/ 50 forms for both; besides current catalog, could not find any publications on the measure				Adult, adolescent, and child versions		13-17, 5-12
Teacher Observation of Classroom Adaptation-R (Werthamer-Larsson et al, 1991)	Measures attentional skills, academic performance, aggressive and disruptive behavior, and social contact; 26 items			Could not find additional data.	Alpha = .92 to .97 across subscales	Confirmed 3 factor solution; Karns (1997): "good concurrent and predictive validity"	Normative sample of 378			Elementary school children

General Symptoms

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Achenbach scales CBCL 4/18 (1991a) YSR (1991b) TRF (1991c)	Measures problem behavior in children over past 6 months; 118-item CBCL; 112-item YSR; each have externalizing and internalizing scales; problem scales include withdrawn, somatic complaints, anxious/ depressed, social problems, thought problems, attention problems, delinquent behavior, aggressive behavior, sex problems;	Self, parent, teacher	Normative sample - 73% white, 16% black, 7% hispanic, 4% other, 81% of sample middle or upper class, no separate norms for ethnic grps or social classes because differences only explained small portion of variance; translated into 33 languages, has been studied in 15 different cultures; YSR similar ethnic background as CBCL; some translations	Written at 5th grade reading level, 15-17 minutes; manual = $29; scoring template = $9/ informant; answer and scoring sheets = $24/25 forms; avail from University of Vermont, Dept. of Psychiatry, Burlington, VT, USA; Cross & McDonald - misrepresents capabilities of children with chronic illness; may be distorted by parents or adolescents who exaggerate or minimize child problems	CBCL Inter-parent = .48 to .98 on subscales; test-retest (1 week) = .70 to .95 on subscales ; mean = .89 However, 1-2 year test-retest = .71 to .75 .67 to .73 for anx scales Cronbach's Alpha >.89 Total, Intern & Extern CBCL and YSR Internal consistency >.70 for most CBCL scale scores; .59 to .86 for YSR scales; Reasonable inter-parent reliability; YSR -1 week test retest Pearson r on subscales = .37 to .91; mean = .79 CBCL:YSR corr. = .14 girls; .29 boys CBCL:YSR corr = .14 to .62 Interrater reliability when used as an interview instrument = .96 YSR 7 week test-retest = .49 - .56; YSR 6 month test-retest = .69	Content validity - 116 of 118items signif assoc w/ child's clinical status -problem items cluster into meaningful scales, problem scales corr .56 to .86 pearson r with Connors; .52 to .88 with Quay-Peterson (construct validity); total and specific problem scales discriminate dift childhood problem grps from comparisons- clinic referred vs. controls, with and without hyperactivity, with and without depression, abused children vs. controls, SED children vs. LD children; 68% referred and 18% non-referred score above clinical cutoff YSR discriminates referred and non-referred youth; status predicted large proportions of scale variance Agreement w/ PL 94-142 system in classifying children w/ significant behavior disorder = 55%; agreement w/ DSM III = 63% YSR I and E scales appropriately correlated with other adolescent self-reports; YSR I corr. .73 with CDI; CBCL anx corr w/ Conners anx = .67; w/ anx RPBC = .78; social desirability related to reduced Externalizing score construct validity not fully assessed; content validity - Salvia & Ysseldyke - items based on clinical research and lit. reviews high intercorrelation of internality/externality - suggests lack of discriminant validity??? (Knight, Virdin, & Roosa, 1994) Construct validity of YSR not demonstrated - Salvia & Ysseldyke	Each sex, 4-11, 12-18; sample of 2,368 children; YSR 11-18; sample of 1315 normals and 1272 clinic sample; All checklists normed on over 5000 individs	Norms for different age groups; self-report for ages 11-18; YSR - test retest Pearson r average = .83 for scales and .91 total score for 15-18 but = .65 or scales and .70 total score for 11-14; 11-14 found on YSR to have .37 1 week test-retest on 2 subscales	Cross & McDonald - has been used successfully to measure the effects of many interventions;	Children and adolescents in treatment; CBCL, ages 4-18; YSR, ages 11-18
ACQ Behavior Checklist (Achenbach, Conners, & Quay, 1983)	Measures emotional/ behavioral problems; contains entire CBCL, 215 items			Too long						

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Behavior Dimensions Rating Scale (Bullock & Wilson, 1989)	Screens for behavior patterns - Aggression, inattentive, socially withdrawn, anxiety; 43 items (7 pt. scale)	Counselors, teachers, parents - must know participant for at least 2 weeks	Hammil, Brown, & Bryant report unacceptable demographics; 67% white, 19% black, 11% Hispanic, 3% other; normed in Northeast, Midwest, South, and West	5-10 min.; kit = $95; $36/25 forms; no training required	Internal consistency = .90 to .95 (subscales), .96 (total). Test-retest .82 to .89 Interrater reliability = .60 to .68 between teachers and teacher assistants	Content validity: experts confirmed subscale content on a 30-item version. Signif differentiates behaviorally disordered (BD) and non-BD children. 74 to 76% correct placement by subscale. Confirmatory factor analysis established adequate fit of hypothesized structure Multitrait Multimethod validation: aggression and inatttentive with good convergent validity, other subscales not conclusive. Scales have not been adequately validated (Vaidya, 1996) - manual makes many unsubstantiated claims	Standardized on 1,942 Canadian subjects; Hammil, Brown, & Bryant rate as having unacceptably low percentage of females; separate norms for 641 youth in correctional facilities; separate norms by gender	Separate norms by grade level to deal with significantly different scores by age		
Behavioral Problems Index (Peterson & Zill, 1986)	Measures behavioral problems in last 3 months; 28 items (yes-no)	Respondent about the child			Alpha = .85 to .89			Different items for children above and below age 12		
Behavior Rating Profile-2nd edition (Brown, & Hammill, 1990);	Evaluates emotional, behavioral, personal, or social adjustment problems at home, school and in peer relationships; 6 separate measures; 3 child scales (Home, School, and Peer); Parent Rating Scale; Teacher Rating Scale; Sociogram	Child rates behavior problems in home, school and with peers (individ or grp); parents; teachers	Hammil, Brown, & Bryant report good demographic mix of normative sample; educational attainment of parents biased toward high end; Spanish lang version; Mexican and US norms for Spanish version	Approx 20 mins; Intro kit = $174; $34/50 forms; available from PRO-ED (USA); sociogram very difficult to score - Salvia & Ysseldyke	Internal consistency = .77 to .98; .76 to .97 for LD and emot disturbed for all reporters; 2 week test retest for all informants = .43 to .96 No interrater reliability data	Hammil, Brown, & Bryant report at least 5 types of empirical validation Content validity - items from literature, examining other scales, and parent input Divergent validity - near zero corrs with achievement and aptitude; 64 of 72 comparisons show correlation of >.35 with BRP - measures such as Walker Problem Behavior Identification Checklist, Quay-Peterson Behavior Problem Checklist, and Vineland Social Maturity Scale (However, Salvia & Ysseldyke say this data is not discussed sufficiently); also, weakest corrs for children with severe emot disturb receiving services; students with social/emotional disturbance and mental retardation differentiated from LD and nondisabled students (unknown if signif diff. - Salvia & Ysseldyke)	Norm referenced subscales - 2,682 children from 26 states for child report (normals, LD, and emot disturbed but SED children excluded for child, teacher, and parent report; 1,948 from 19 states for parent report; 1,452 from 26 states for teacher Norms gathered and integrated from 1978, 1983, 1989	Test-retest reliability lowest for grades 1 and 2 = .43 to .94		

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Bristol Social Adjustment Guides (Stott, 1969)	Measures maladjustment in the school setting; 146 items	Teacher, social worker, or other adult familiar with child		Approx. 15 mins; avail from Hodder & Stoughton (UK)	Alpha = .68					
Burks' Behavior Rating Scales (Burks, 1977)	To identify the nature and severity of pathological symptoms; 110 items (5 pt. scale)	Parent or teacher		20 min.; has intervention recommendations for problem areas; kit = $145; $28.50/25 forms	Internal consistency <.80; test-retest <.80;	Hammil, Brown, & Bryant report that some subscales have been empirically validated while others have not	No normative data			
Devereux Scales of Mental Disorders (Naglieri, LeBuffe, & Pfeiffer, 1994)	Identifies behavioral or emotional problems based on DSM-IV criteria; 110-item child form; 111-item adolescent form; 5-pt. frequency scale	Any adult who has known child for at least 4 weeks (parent, teacher, or other adult)		Kit = $159.50; $49/25 forms; 15 mins (admin), 10 mins (scoring); requires PhD to purchase; 6th grade reading level required; Psychological Corporation, USA; instructions for determining statistically and clinically meaningful change are provided (LeBuffe & Pfeiffer, 1996)	Excellent scale reliabilities- No data given	Content validity based on DSM-IV and mapping of specific items to corresponding DSM-IV criteria	National norms - 3000 cases by gender and by teacher or parent informant	Child version (5-12 years of age), adolescent version (13-18 years of age)		
Eyberg Child Behavior Inventory (ECBI; Eyberg & Ross, 1978)	Identifies behavioral problems and measures problem severity; 36 items				Test - retest = .86; Internal consistency =.98		512 children			
Louisville Behavioral Checklist (LBCL; Miller, 1988)	Screens for problem behaviors, four forms; 164 items (true-false)	Parent		20-30 mins; 10 min. to score; 10th grade reading level required; kit = $195; $19.50/100 forms; No additional data could be found on this measure	"Reliability levels...are variable" (Zimmerman, 1996, p. 58)	"Validity is questionable" and adolescent version has "inadequate validity research" (Zimmerman, 1996, p. 58),	Norms lacking according to Zimmerman (1996)			

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Ontario Child Health Study scales (OCHS; Boyle, Offord, Racine, Fleming, Szatmari, & Sanford, 1993); parent and teacher versions	Behavior/symptom checklist for the determination of DSM-III-R-related disorders; scored 0, 1, or 2, indicating “never or not true,” “sometimes or somewhat true,” and “often or very true,” time frame within the past 6 months	Parent or teacher			Parent: test-retest (average of 48 days) using Cohen’s Kappa ranged from .27-.61 for 6 major diagnoses (*M* = .42)	Parent: used Cohen’s Kappa to measure convergent validity with the DICA-R (Reich & Welner, 1988). Agreement on cases for 6 major diagnoses ranged from .05-.52 (*M* = .34). Positive predictive values (% accurate) for accessing help ranged from 70-92 (*M* = 77). Area under the curve (sensitivity plotted against specificity) ranged from 62-79 (*M* = 72). Relative odds between classifications of disorder and impaired social functioning (ISF) was significant in 4 of 6 diagnoses (parent report of ISF) and significant in 3 of 6 diagnoses (teacher report of ISF). Relative odds for poor school performance were significant in 5 of 6 diagnoses (parent report) and 2 of 6 diagnoses (teacher report). Hypotheses for significant relative odds derived from forward stepwise regressions between classification of disorders and selected correlates (discipline problems, developmental delay, and depressed mood) were correct on 4 of 6 diagnoses (convergent and discriminant validities).				
Parent Daily Report (Chamberlain 1980); parent and adolescent versions	Measures problem behavior occurrence over a 24 hour period; factors include: overt aggression; verbal negativism; non-compliance; hyperactivity; depression; covert antisocial behaviors; 31 items; yes/no format	Parent observation checklist; adolescent			Interparent reliability reported in source ref.; Test-retest = .85 to .98; Alpha of subscales = .23, .54, .76, .81, and .83. Interrater reliability = .85 - .98.				Reported as sensitive to change (Chamberlain, Ray, & Moore, 1996)	

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Pediatric Symptom Checklist (PSC; Murphy & Jellinek, 1985)	Measures parent's view of children's psychosocial functioning; designed as a screening tool, not a diagnostic tool; 35 items; rates "never," "sometimes," or "often"; scores up to 27 within normal range for children ages 6-12	Parent	Spanish version available; case rates associated with SES (Pearson r = -.12); higher specificity in lower-SES samples than middle class (100% vs. 68%), and higher sensitivity in middle class (95% vs. 80%); economically disadvantaged significantly more likely to test positively	5 mins; parent completes a 1-page, 2-sided form in a waiting room or oral administration; "well-liked by parents, praised by pediatricians, and accepted by clinic staff" (Jellinek et al., 1995, p. 742); 90% completion rate in clinic	Internal consistency: Spanish version *r* = .91; English version *r* = .89 to .94; case rate higher for oral administration (14.1% vs. 7.4%) with Mexican-American pediatric clinic population; Percent agreement = 63-92%, kappa = .52-.82. Test-retest (1 week) = .85 and .86 in 2 samples; agreement (1 week) on caseness = 91%, kappa = .69; Cronbach's alpha = .86 4 week test-retest = .86 36 day test retest kappa = .68	Content validity: shortened and revised Washington symptom checklist. Criterion validity: 69% of children referred to mental health services case-defined compared to 4% of non-referred children. Corr .64 w/ CBCL; .63 w/ CGAS. Sensitivity = 87-95%; specificity = 68-100%; using the psychologist's rating of the CGAS as the standard for caseness, sensitivity = 42% and specificity = 87%; kappa = .60; using pediatricians' ratings of caseness, sensitivity = 43% and specificity = 93%. In one study, on ratings of caseness compared to clinicians - kappa = .75 Agreement with CGAS caseness kappa = .82 W/ DICA - sensitivity = .87; specificity = .89; kappa = .74 Signif differentiated children high and low in stress Signif differentiated high and low FAPGAR scores Kappa agreement with CBCL for caseness = .52 at selected cutpoint of 28 Agreement w/ counselor ratings of student dysfunction = 83%. Distinguised between cases (*M* = 32.2) and non-cases (*M* = 11.7).	Norms available	Different cutoff score for younger children		
Revised Behavior Problem Checklist (RBPC; Quay, 1983)	Checklist of behavior problems; scales: conduct disorder, socialized aggression, attention problems, anxiety, psychotic behavior, motor tension excess; 89 items; 3 pt. scale; Cross & McDonald -criticize ambiguous zero score	Caregiver or teacher	Examined old version for African-American and native American sample but no work published on revised measure; measure tested with populations of lower, middle, and upper classes; 3:1 male to female ratio in norm sample; 89% white, 13% black .30 to.31 corr w/ SES; .14 to .16 corr w/ race; norms not based on representative national sample	20 mins; scoring and profiling 10 mins; requires clinician with some training in administration and scoring of psychological tests; intro kit = $69; $47/ 25 answer and scoring sheets	Cronbach's Alpha of subscores from .70 to .97; Interrater reliability ranged from .52 to .93; teachers .52 to .85; parent pairs .55 to .93 2 month test-retest .49 to .79 5 mo. Test-retest on anx scale = .72 Test-retest = .49 to .83 2 year test-retest = .82 Interrater test retest w/ teachers 7 to 17 mos = .32	Cross & McDonald - .43 to .92 correlations of subscales with CBCL subscales .36 to .72 correlation with scales of sociometric pupil evaluation inventory; .43 to .49 corr w/ teacher rated probs; -.49 to -.36 corr w/ teacher rated competencies .00 correlation of motor tension scale with observations of motor activity; discriminated btwn clinical and nonclinical samples, between behavior disordered and non BD kids; 85% correct identification of BD kids anx scales signif diff. for clinical vs. normal; anx dx kids from symptom free controls 7 of 21 significant correlations with peer ratings Significant correlation with achievement and aptitude significantly discriminated inpatient, outpatient, and LD children from community sample; 86% correct classification for boys; 91 % girls, 23% false negatives	Teacher report norms for K-12; no analysis by gender; 4 clinical samples from 5 to 23 y.o.	No analysis by age	No studies report using this measure for program evaluation	

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Rutter's Parent Questionnaire (Rutter, Tizard, & Whitmore, 1970)	Assesses emotional and behavioral symptoms; 31 items (3 pt. scale) Also has teacher version	Parent	Has been translated into Finish	Old version of Modified Rutter A2 Parent Questionnairre (Goodman, 1994)		.25 corr. w/ CGAS; .49 at 5 months post treatment				
Vermont System for Tracking Client Progress (Burchard & Bruns, 1993)	Measures behavioral, emotional, life event, and educational outcomes considered indicators of risk of out of home placement; consists of Daily, Weekly, Monthly, and Quarterly Adjustment Indicator Checklists; 16 negative and 6 positive behaviors; also includes restrictiveness of Living Environments Scale;	Service Providers or primary caregivers	Only studied in Vermont with predominantly white males		Cronbach's Alpha = .77 to .85 for negative behaviors;	Cross & McDonald - Discriminated between wraparound and comparison group; >.60 Pearson R to CBCL for problems: .50 Pearson R with restrictiveness and .50 to .59 with one year case management costs; Compliance -.54 with restrictiveness and -.70 with cost factor analysis provide some evidence of validity - however, factors not found to be clear across analyses, more testing needed may be subject to case manager biases in rating improvement		Does not specify what ages that it is appropriate for	Client group demonstrated significantly greater improvement vs. comparison group	
Walker Problem Behavior Identification Checklist (Walker, 1976)	50 items, subscales: acting out, withdrawn, distractability, disturbed peer relations, immaturity		Only normed in Oregon and Washington		Split half reliability = .98; test-retest = .43 to .88 Inter item corrs = 0 to .83	Content validity - teachers and experts involved in item selection process; Signif predicts Behavior disordered children; child withdrawal; achievement, classroom observations of behavior, deviance Exploratory factor analysis supports instruments theoretical basis	Normative sample = 1855	Separate norms for ages 2-6 and grades 1-3	In 6 separate studies significantly demonstrated intervention effects	

Goals

Adolescent Goal Attainment Scale (Maher & Barbrack, 1984)	Measures and evaluates goal attainment; one to five goals, 5 pt. scale of anticipated outcome)	Student; counselor		Not used in 1990s	Weighted Kappa agreement btwn student and counselor = .42 to .82	Correlation =.48 btwn goal attainment scores and teacher ratings of satisfaction with students				
Family Goal Recording (Fleuridas, Rosenthal, Leigh & Leigh, 1990)	Method to operationalize and measure change related to target complaints			Procedure to use by therapist during session to generate goals	Agreement between rater and therapy teams Kappa = .78	Content validity - adaptation of GAS; -.62 corr w/ mother's dissatisfaction with children but .22 for father; corr. w/ mother's FAD = -.43 but only .16 for father, -.43 for identified child; corr w/ father's global distress w/ marriage = -.68 but only -.26 for mothers				

Maltreatment

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Adolescent Abuse Inventory (Sebes, 1983)	Measures parental attitudes toward maltreatment and likelihood that they would act abusively given provocative adolescent behavior; 26 items (hypothetical situations)				Alpha = .607 to .877 (mothers); .644 to .881 (fathers)					Families with adolescents
Checklist for Child Abuse Evaluation (Petty, 1990)	For investigating and evaluating possible abuse or neglect; 264 items	Child and witness report; clinician judgments		Requires some training in psychological test administration and scoring; Introductory kit = $56; $89/ 25 checklists; Besides being in recent catalog, cannot find in any publications						Children and adolescents
Child Abuse Potential Inventory (Milner, Gold, & Wimberly, 1986)	A screening device for differentiating physical abusers from nonabusers; 160 items (77 items in physical abuse scale); agree-disagree format; has 3 validity scales.	Parents/ caregivers.	Normed group w/ average of 14 years of education. Tested w/ African-American adolescents. Translated and tested in Croatia and Greece; Greeks scored signif higher than Americans; translated and tested in Spanish. Significant, but low correlations in some studies (income and education), although most associations were non-significant; in a large sample, significant corrs found, but each demographic factor accounted for 2% variance or less; In Greece - low SES, nonmarried parents and rural parents scored signif higher	12-20 mins; 3rd grade reading level.	Split-half reliability = .96 to .98. KR-20 reliabilities = .92 to .95. Internal consistency = .74 to .98 across several samples and studies. As low as .24 on one subscale of Greek version 1 day test-retest = .91; 1 week test-retest = .90; 3 mo. test-retest = .75.	Content validity - from comprehensive lit review of child abuse and neglect; many items eliminated that did not significantly discriminate abusing and nonabusing parents; a later study found 77 of 160 items signif discriminated. Pos corr. with self-reported hx of maltreatment; corr with high level of anger during parent-child interactions Sample of maltreating families (abuse and neglect) receiving treatment signif higher score than norms or control groups (several studies); also differentiates spouse abusers from nonabusers Correct classification of abuse and neglect group = 77%; of controls w/ no reports of abuse or neglect = 89%; average = 88 to 93% across studies. Abused adolescents and witnesses of violence scored significantly higher than nonabused. In a small sample of referred, low-SES families:.55 corr with negative parenting style, .79 with parent global severity, .32 with CBCL int, .45 with CBCL ext, and .61 with CBCL total; .29 with TRF scores; -.11 with belief in corporal punishment; -.17 with parent problem solving effectiveness and -.04 with mean no. of problem solutions .27 corr. w/ report of child abuse In one sample - 25% spouse abusers and 31% nonabusers with elevated lie scale. Greek version factor analyzed - items did not all go to factors from original factor analysis; another factor analysis found different number of factors than original	Normed on 836 adults	Adolescents score significantly higher than adults	Sensitive to changes in abuse potential following involvement in abuse prevention services	Parents (adolescents and adults)

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Childhood Trauma Questionnaire (Bernstein et al., 1994)	28 items; screens for hx of child abuse and neglect; has a minimization/denial scale			5 min.; 1 to 2 min. to score; includes suggestions for clinician use; $75/intro kit; $30/25 forms; almost every study has used this with adults; very little use with adolescents or children; retrospective measure	High test-retest?	Factor analyses support the five scales; good convergent and divergent validity with trauma hx determined by other measures; sensitive at identifying adolescents with confirmed hx of abuse or neglect	Normed on 2200 individs from 7 dift. Clinical and community samples			12 and older
Children's Impact of Traumatic Events Scale - R (Wolfe et al., 1991)	To assess the impact of sexual abuse; 78 items	Standardized interview of child		20 to 40 min. too long						8 to 16
Child Report of Treatment Issue Resolution (Nelson-Gardell, 1997)	Assesses effectiveness of therapy intervention for sexual abuse victims; 38 items	Self-report			Alpha = .93	Content validity thru expert ratings of individual item relevance; Corr w/ Trauma Symptom Checklist for Children = -.74; corr with CBCL = -.41 to -.34; Poor discriminant validity as .45 corr w/ Children's Social Desirability Questionnaire; 4 factors accounting for 47% of variance				10 to 17
Child Sexual Behavior Inventory (Friedrich, 1990)	Assesses sexual behaviors observed in the last 6 months. To evaluate children who have been or may have been sexually abused; 38 items (4 pt. scale)	Female primary caregivers		10-13 min to admin and score; 5th grade reading level; $69.00 for intro kit; $36/25 kits; requires advanced training in psychological testing and interpretation			Has age and gender norms - normed on 1,114 child community sample and 512 children from child abuse centers			Children 2-12
History of Victimization Form (Gentile, 1988)	Checklist of sexual experiences	From clinician interviews with children				2 factors accounted for 65% of total variance				
Sexual Abuse Fear Evaluation Subscale (Wolfe & Wolfe, 1986)	27 item subscale of Fear Survey Schedule for Children; (3 pt. scale) - measures sex associated fears and interpersonal discomfort				Alpha = .80 to .81 for factors					

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Trauma Symptom Checklist (Briere, & Runtz, 1989)	Evaluates acute and chronic posttraumatic symptoms ; 54 items (4 pt. scale); has 2 validity of response scales; alternate 44 item version without sexual issues	Self-report (individual or group administration)	Urban and suburban normative sample.	15-20 min.; easy hand scoring; requires clinician with some training in psychological test administration and scoring - minimum masters degree; Introductory kit = $79; $110/ 100 booklets; $13/25 profile forms; has computerized scoring program = $159	Alpha = .77 to .89 for subscales	Corr w/ CBCL = .32 (Int = .37 and Ext = .19); corr w/ YSR = .80 w/ Int and .77 w/ Ext Corr w/ Child Report of Treatment Issue Resolution = -.74	Standardized on over 3000 inner city, urban, and suburban children and adolescents; data from trauma and abuse centers; male and female profile forms			Children (8-16) who have experienced traumatic events (abuse, loss, disasters, witnessing violence)

Multidimensional Measures

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Adolescent Treatment Outcomes Module (Robbins et al., 1997)	Assesses symptom severity, functioning at home, school functioning, troubles with rules and laws, leisure functioning, friendships, dysfunctional peers, family burden, and satisfaction; 68 items	Child self-report and caregiver/ parent		4th grade reading level; 28 min. for parent, 25 min for child, 7 min. for clinician	Test-retest = .64 to .86 across scales; Alpha = .60 to .94 across scales and informants	Content validity - items from review of literature and panel of experts suggesting what is relevant for understanding how treatment affects outcomes; Differentiated DISC-dx of conduct disorder (sens = .77, spec = .92; kappa = .65); ODD - (sens = .60; spec =.88; kappa = .47); anxiety (sens .82; spec .74; kappa .52); depression (sens .88 spec .61; kappa .50) Did not differentiate ADHD (sens .53; spec. .61; kappa .37) Spearman corr of focal problem severity w/ CBCL/YSR internalizing = .51 parent; .13 adol Spearman corr of symptom severity w/ CBCL/YSR internalizing = .69 parent; .43 adol; w/ CHQ behavior probs = -.74 par; -.23 adol Spearman corr of func at home w/ CIS = .47 parent; .42 adol; w/ CHQ subscale = -.45 par; -.37 adol Scorr school w/ CIS = .47 par; .40 adol Scorr trouble w/ CIS = .33 par; .44 adol Scorr leisure w/ CIS = .37 par; .44 adol Scorr friendships w/ CIS = .43 parent; .44 adol Scorr burden w/ CHQ scale = -.53; w/ another CHQ scale = -.55			Corrs w/ changes in other outcome measures; parent .10 to .72; adol .07 to .43; over 6 months post treatment but no control group;	11-19 with moderate to severe impairment in functioning

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Behavior Assessment System for Children (BASC; Reynolds & Kamphaus, 1992); consists of a Teacher Rating Scale (TRS), a Parent Rating Scale (PRS), a Self-Report of Personality (SRP), a Structured Developmental History Inventory, and a Student Observation System (classroom)	Provides composite scores for behavioral symptoms, externalizing, internalizing, school problems, and adaptive skills; self-report composite scores include emotional symptoms, school maladjustment, clinical maladjustment, and personal adjustment; hand or computer scored; 109, 148, and 138 items; 4-pt. scale; has validity of report scores	Parent, child, or teacher	Norms were attempted to balance with US census including exceptional children; Midwest overrepresented and North west underrepresented; gender or ethnicity biased items dropped during development; different norms reported for males and females	10 to 20 mins for parent and teacher; 30 mins for child; SOS requires 15 minutes of observation. This measure has multiple ratings, two of which are not suitable due to teacher informant and classroom observations. Manual $50; $20/25 forms. Available from AGS, Circle Pines, MN, USA.	Internal consistency, teacher >.80 for all scales across all age groups; internal consistency, across ages parent = .56 to .94; internal consistency, child across ages = .54 to .97 (mean=.80); 2 to 8 week test-retest = .82 to .91 for teachers across ages; 2 to 8 week test-retest = .70 to .88 for parents across ages; 7 month test-retest = .69. Child 1 month test-retest = .76 to .86 Teacher interrater = .29-.93 (median = .83) for preschool and child; none for adol. Parent inter rater = .46 to .67 across ages;	Content validity: items based on viewing other measures, literature, and clinical experience; data driven determination of items in phases; scales from factor analysis. TRF Externalizing scale corr .90 with 5 other measures; internalizing btwn .73 to .81; only low to mod corrs across scales with Behavior rating Profile; Parent report corr with Connors Parent Rating Scale - .67 to .78 for externalizing and .45 to .51 for internalizing; low to moderate corr with Behavior Rating Profile High corr of self report with MMPI. Does not adequately discriminate btwn referred and non-referred samples (Salvia & Ysseldyke). Corr of SRP with YSR “is limited” (Flanagan, 1994, p. 182).	Community normed between 1988 and 1991; 116 testing sites; 2,401 teachers, 3,483 parents, and 9,861 children. Clinical norms in 36 sites; 693 teachers, 401 parents, 411 children. Gender-based norms	3 different versions 4-5, 6-11, 12-18; for parent and teacher; child version: 8-11 and 12-18 children with behavioral and emot problems appropriately represented at each grade level All reliabilities examined across ages Low to moderate corrs increasing with age btwn parent and teacher		Ages 4-28
Child Behavior Rating Scale (Kilpatrick, 1975)	Measures 16 social and personal behavior areas (8 prosocial, 8 problem areas), 146 items, 4 pt. frequency scale	Parent or child care worker (residential care worker)			Test-retest (10 days) ranged from .73-.97); parental agreement *r*s significant on 15 of 16 subscales (.57-.79; *M* = .65); child care worker agreement *r*s significant on 12 of 16 subscales (.40-.62; *M* = .45)					
The Health of the Nation Outcome Scales for Children and Adolescents - Version 6.0; (Gowers, Whitton & Harrington, 1998)	Measures symptoms, functional disability over past 2 weeks; 13 items (5 pt. scale)	Clinician rating				Corr. with CGAS = -.565; Corr. with Child Health Related Quality of Life-parent = .445; Corr. with Child Health Related Quality of Life-child = .380; Corr with Strengths and Difficulties = .437; corr with Harter Self-esteem = -.375; Corr. with Behaviour Checklist = .338				

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Life Stressors and Social Resources Inventory-Youth (Moos, Fenn, Billings & Moos, 1989)	Measures ongoing life stressors and social resources	Self; individ or grp		30-60 min.	Internal consistency = .66-.92 for stressor scales; .78-.93 for social resources scales		Normed on 400 youth			Adult and youth versions (12-18)
Matson Evaluation of Social Skills with Youngsters (MESSY; J. L. Matson, Rotatori, & Helsel, 1983)	Measures impulsiveness/ inappropriate assertiveness and appropriate social skills; 64 items (5 pt. scale)	Self			2 week test-retest = .50 Alpha = .87	Significantly differentiated autistic from nonautistic children 5 factors from factor analysis Corr. w/ Vineland = .28 (not signif) Corr. w/ 3 depression measures = -.39 to -.47	Normed on 322 children			4 to 18
Mental Status Checklist (Dougherty & Schinka, 1988)	Assesses presenting problems, recreation and reinforcers, personal information, family/peer relationships, physical/ behavioral obs, devt. Status, academic performance and attitudes, health and habits, legal issues/aggressive behavior, impressions, recommendations; 153 items for child; 174 items for adol	Clinician checklist		No training requirements; 20 min. ; $37/25 forms	Split half reliability of Parent Prosocial score = .85; Alpha = .82 Corr parent prosocial w/ teacher prosocial = .34	Sensitivity for caseness against psychiatric assessment = 71%; specificity = 70% using standard cutoff; Corrs of subscales w/ interview derived same area of psychopathology = .48 to .65; corrs in MTMM w/ off diagonals - discrim valid .03 to .44 For children w/ normal intelligence Corrs of subscales w/ interview derived same area of psychopathology = .19 to .54; corrs in MTMM w/ off diagonals - discrim valid -.11 to .27 For children w/ learning difficulties Factor analysis 15 items w/ eigen values >1 but only kept the 6 that made clinical sense Parent prosocial corr w/ Total deviance, conduct and hyperactivity score = -.18 to -.27 Parent prosocial rating corr w/ IQ = .12 for IQ >70 but .61 for IQ <70		Adult, adolescent, and child versions with age approp developmental issues		13-17, 5-12
Modified Rutter A2 Parent Questionnairre (Goodman, 1994)	Measures emotional and behavioral difficulties and also prosocial behavior, positive traits,; 50 items			Could not find any data on this measure						

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Ohio Youth Problems, Functioning, and Satisfaction Scales (Ogles, Davis & Lunnen, 1998)	Designed to be sensitive to changes in client strengths and weaknesses that youth make during intervention; domains:- some domains vary by informant. 44 problem items, 20 functioning items, 1 satisfaction item, 3 inclusion items, 4 hopefulness and wellbeing items, 2 family functioning items (varying points likert scales)	Parent, youth and case manager self report	Normative sample from southeast Ohio; does not have separate gender norms	Requires minimal training; 8-10 minutes	Internal consistency = .76 to .97 across subscales for parent; .74 to .96 for youth; .44 to .94 for case managers	Content validity - reveiew of research studies, conceptual development process to determine areas relevant to clinical outcomes assessment; stake holder survey conducted to identify areas deemed most important; items from DSM-IV, CMHC list of common presenting problems, survey, review of instruments, and consulation with service providers. Problems-P corr w/ CBCL = .89; .77 for functioning-P Problems-Y corr w/ YSR=.82; .46 for functioning-Y Problems-CM corr w/ Progress Evaluation scales = -.47; .38 for functioning-CM; Change in problems-CM corr -.54 w/ changes in PES; Change in functioning-CM corr - .56 with changes in PES 4 of 5 groups (paired comparisions) were found to have significantly different means on symptoms, functioning or both.	Normative clinical and community sample - clinical cutoffs based on t-scores -	Does not have separate norms by age		
Paddington Complexity Scale (Hunter, Higginson, & Garralda, 1996)	Measures psychiatric case complexity; 16 items	Rater				Corr. with CGAS = -.630; Corr. with HoNOSCA = .522; Corr. with Child Health Related Quality of Life-parent = .162; Corr. with Child Health Related Quality of Life-child = .175; Corr with Strengths and Difficulties = -.431; corr with Harter Self-esteem = -.142; Corr. with Behaviour Checklist = .472				
Questionnaire developed for specific study (Slee & Rigby, 1993)	Measures tendency to be bullied, to bully others, and to act in prosocial manner, also questions about friendships, school affect, and how safe child feels at school					3 factors; 47% of variance explained tendency to be victimized items correlated -.21 to -.41 with # of friends; -.15 to -.37 w/ popularity w/ peers; -.27 to -.47 w/ perceived safety at school items (all items in same questionnaire)				7 to 13

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Scales of Independent Behavior (Bruininks Woodcock, Weatherman, & Hill, 1984)	Measures independence, maladaptive behavior, living skills, etc.			45 min.						0-3 to 44-11
Service Utilization and Risk Factors interview (Goodman et al., 1998)	Measures envt. and family risk factors and social competence			41 min. for youth; 68 min. for adult respondent						
Social Adjustment Inventory for Children and Adolescents (Gammon, John, Prusoff, & Weissman, 1982)	Systematic assessment of children's adaptive functioning in school, spare time activities, with peers, siblings, and parents; 77 items - 35 competence items and 42 problem behavior items (4 pt. scale)	Semistructured interview of children or parents	Not tested with different ethnic groups	To be administered by interviewer knowledgable in child devt. and with clinical or testing experience; interview has been found acceptable to parents and children; 30 min.; takes 4 hours of training to minimum of master's level of interviewer to get 95 % agreement on corated inperson and video ratings	Corr. btwn parent and child report = .10 to .72 on various subscales; Interrater agreement kappa = .85 to 1.00	Factor analysis - parent - 3 factors - 24, 15, and 17% of variance; child - 17, 17, and 17% of variance Mother and child report of compentence corrs w/ WISC vocab .24 to .41; corr w/ probs = .04 to .10; competence w/ PPVT = .23 to .50; w/ probs = .03 to .08 discrim. valid Mother and child report of competence corrs w/WISC Block Design = -.11 to .30 (not good); corr w/ probs = .05 to .20; Corr w/ CBCL competence = .52; w/ CBCL total probs = .64 Competence corr w/ C-GAS = .20 to .39; Probs corr w/ C-GAS = .41 to .56 Significantly differentiated children with no diagnosis from children with a diagnosis for mother but not child report Several studies find for adult version: children w/ a dx with higher score than normal controls; also, children with comorbid psychiatric disorders found higher score than ADHD only children; child version does not discriminate children w/ dx from controls	Not normed	Has an optional section for use with older children		6 to 18

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Social Skills Rating System (Gresham & Elliot, 1990); elementary and secondary levels	Measures social skills in several areas, problem behaviors (externalizing and internalizing; hyperactivity at the primary level), and academic competence; 55 items	Teacher, parent, student	Hammil, Brown, & Bryant report inadequate demographics for normative sample	20-25 mins; American Guidance Service, USA, 1990	Internal consistency at least .80 for most teacher subscales; Most child and parent subscales <.80; Alpha = .87 for Social Skills scale; >.74 for 3 problem scales Test-retest = .87 for Social Skills and .48 to .72 for Problem scales Parent version signif corr with teacher and peer versions??	Hammil, Brown, & Bryant report not validated	Normed on at least 750 subjects with 75 subjects per age			Preschool through grade 12; child report 3rd grade to 12th grade
Strengths and Difficulties Questionnaire (Goodman, 1997)	A behavioral screening questionnaire - measures strengths and difficulties with dimensions of conduct problems, emotional symptoms, hyperactivity, peer relationships, and prosocial behavior; 25 items (3 point scale) Also has parent, teacher, and self-report specific impact supplements - burden items	Teacher, parent, and youths (11 and up)		Significantly more mothers preferred SDQ over the CBCL	Parent-teacher corrs = .37 to .62	Content validity - initial item choice guided by factor loadings and frequency distributions from expanded Rutter parent questionnaire - items modified based on informal trials and advice from colleagues Sensitivity .87 for discriminating dental and psychiatric patients - same as ROC for Rutter Questionnaire. Specificity = .97 Self-report version signif discrim community and clinical samples Corr. of subscales w/ Rutter = .78 to .88 for problem scores for parent; .87 to .92 for teacher SDQ burden corr = .66 w/ standardized interview rating of burden Corr. SDQ domains w/ CBCL domains = .53 to .87 Corr. w/ problem areas inteview = .41 to .52				4 to 16
Timberlawn Child Functioning Scale (Dimperio, Blotcky, Gossett, & Doyle, 1986)	Has 14 scales related to cognitive, affective, and social functioning, estimates average adaptive level for the last year; Likert scale format.				Corr. teacher w/ primary mental health worker = .62	Corr. w/ child externalizing probs = -.33				

Quality of Life

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Child Health and Illness Questionnaire (CHIP; Starfield et al., 1993)	Assesses adol health status in 6 domains: activity, comfort, well-being, disorders, achievement, and resilience; 5-pt. scale (0 = healthy; 4 = least healthy)	Adol	Slightly more blacks than whites and more girls than boys in the validation sample	45 minutes-too long (shorter version forthcoming, or may be reported in a later publication)		Content: used existing literature for instruments that tapped relevant domains; used focus groups of children, adolescents, and parents to define "health"; sent sample items to >50 experts who indicated gaps in domains; sent items to 10 experts who sorted them into appropriate domains; pilot tested the instrument for feasibility, readability, and completeness with 9 adolescents.				
Child Well-Being Scales (Magura & Moses, 1986)	43 scales measuring child abuse and neglect, quality of parenting, provision of adequate resources, provision for physical and mental care, quality of family interactions, child abnormalities, child educational matters, and parental recognition and response to child/family problems					Content: lack of clear definition or conceptual model guiding development (Seaberg, 1988) Factorial: the 3 factors found were broader than those proposed by author; 15 scales independent of any factor (Seaberg, 1988).				
Hopelessness Scale for Children (Kazdin & Petti, 1982)	Measures hopelessness; 17 items (true, not true)	Child - items read to child		1st to 2nd grade reading level	Item-total correlation = .19 to .71; mean = .44 alpha = .45 to .97; split-half = .70 - .96 6 week Pearson test-retest = .52	Corrs .49 to .58 with Children's Depression Inventory; .22 with Depression Symptom Checklist; corr w/ Bellevue Index of Depression = .44; corr w/ Self-Esteem Inventory= -.61; corr w/ MESSY (social skills) = -.39. Corr w/ CBCL social competence = -.03 to -.12; corr w/ CBCL probs = .10. 2-factor structure accounted for 29 to 78% and 8.9 to 22% of variance.		Alpha is .16 to .20 lower for young children and much weaker factor structure with adolescents		Children

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Perceived Life Satisfaction Scale (PLSS; Smith, Adelman, Nelson, Taylor, & Phares, 1987)	Measure's life satisfaction with 5 subscales: material and physical well-being, relationships, environment, personal development and fulfillment, and recreation and entertainment; 19 items; 6 pt. scale				Test-retest = .85. Internal consistency = .70 to .80; .76 to .81 across factors in 4-factor solution.	Factorial: 4-factor solution was the most interpretable and accounted for 57% of the variance; the solution was considered "complex and difficult to interpret" (Huebner & Dew, 1993, p. 347). Significantly correlated "with a measure of depression" and demonstrated "the ability to discriminate adolescents in regular classrooms from adolescents referred for mental health services" (Dew & Huebner, 1994, p. 186) Correlated .58 w/ Students' Life Satisfaction Scale				Adolescents
Quality of Life Profile: Adolescent version (Raphael, Rukholm, Brown, Hill-Bailey, & Donato, 1996)	Measures quality of life - extent person enjoys the important possibilities of his/her life; 54 items (5 pt. scale)		One subscale corr w/ SES = .22		Alpha = .94; .67 to .87 for subscales	Content: developed from group meetings with high school students and guidance counselors and review of literature; adolescents used to verify relevance of content and for omissions Corr w/ Self-esteem = .56, Life satisfaction = .51; social support = .51, life chances = .45 Corr w/ tobacco use = -.02; alcohol use = -.21		Corr w/ age = -.16		

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Students' Life Satisfaction Scale (SLSS; Huebner, 1991)	Measures global life satisfaction; 7 items; 4 pt. scale	Self-report		No differences across gender; SLSS correlated significantly with SES (.34); African-American students reported lower life satisfaction than Caucasian students. Internal consistency = .85 for black and white students. 1 factor emerged with both black and white students; degree of factorial similarity between black and white students (using a coefficient of congruence) = .99. Pothoff's test for bias revealed no significant difference between the concurrent validity coefficients for black and white students. Huebner (1994) recommended using the SLSS for research purposes only.	Test-retest (1-2 weeks) for grades 4-8 = .74. Internal consistency = .82 in grades 3-7. Internal consistency = .86 in grades 8-12; item-total corrs = .49 to .73. Internal consistency = .85 for black and white students.	Factorial: has a 1-factor structure (Huebner, 1991a); 1 factor emerged and accounted for 55% of the total SLSS variance (Grades 8-12); factor loadings ranged .61-.83; 1 factor emerged with both black and white students; all items loaded onto same factor except one item even when mixed with 20 positive and negative affect items. Construct: available in Huebner (1991a) and Huebner (1991b); correlated .58 with the Perceived Life Satisfaction Scale, -.52 with the Nowicki-Strickland Locus of Control Scale-Short Form (greater internal locus of control), and .52 with the General-Self subscale of the Self-Description Questionnaire-II (more positive self-image); corr .57 with the Happiness and Satisfaction subscale of the Piers-Harris Self-Concept Scale (1984); a regression revealed that the Happiness and Satisfaction subscale accounted for 32% of the variance on the SLSS. Corr .48 with a 1-item parent rating of child's global life satisfaction; corr with parent global ratings = .41 (black students) and .52 (white students). Divergent: corr w/ Pier-Harris Self-Concept Scale subscales (not Happiness and Satisfaction) = .29 to .40.		Did not significantly correlate with age (-.09) or grade (-.04)		Grades 3-12
Quality of School Life Questionnaire (Epstein & McPartland, 1978)	Assesses students' views of school life; 40 items	Child report			Cronbach's Alpha = .89 KR for Satisfact. w/ school scale = .79 for Commitment to classwork = .80	Satisfact and Commitment subscales corr = .79				

Safety

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Feelings of Safety items (Schwab-Stone et al., 1995)	Measures beliefs about personal safety at home, in school, in the neighborhood, and on the way to school; 4 items; 4 pt. scale	Adolescent self-report	Ethnic minority students reported feeling more unsafe than Caucasian students); students of low SES reported feeling unsafe in more contexts than students of higher SES			Corr .19 with exposure to violence		8th and 10th graders reported feeling more unsafe		Used with children and adolescents in grades 6, 8, and 10

Satisfaction

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Client Satisfaction Questionnaire-8 (CSQ-8; Larsen, Attkisson, Hargreaves, & Nguyen, 1979)	Assesses global satisfaction with services received by child over past 6 months; 8 items; several instruments in the literature with the same name; 4 pt. scale	Caregivers	No data available	5 mins; very low cost	None	Cross & McDonald -Face valid; validity not studied; subject to positive bias does not look at components of care				Children of any age who receive services
Consumer Satisfaction Questionnaire (author unknown)	Assesses child improvement, usefulness of what learned in parent training, ease of learning treatment components; 35 items; 7 pt. scale.				Not studied	Not studied				
Satisfaction Scales (Brannan, Sonnichsen, & Heflinger, 1996)	Satisfaction with specific service settings across 8 content areas: access and convenience, child's treatment, parent services, family services, relationship with therapist, staff responsiveness, financial charges, and discharge/transition; scales for each of 8 service areas: 23 to 51 items; 5 pt. scale	Parent or adolescent self-report	Tested with predominantly white military families; no significant correlations with race, gender, parent age, education or income	10 mins for each service setting questionnaire	Alpha for parent versions (7 service settings) ranged .76-.92 (outpatient), .79-.94 (inpatient/ RTC), .74-.91 (group home), . 77-.93 (in-home), .57-.91 (day treatment), .84-.96 (therapeutic home), and .74-.89 (case management), global satisfaction across service settings ranged .94-.98. Parent intake assessment Alpha ranged .77-.94 across 8 content areas. Alpha for adolescent versions (3 settings) ranged .61-.63 (low end) to .91 (high end) across content areas. Adolescent intake assessment Alpha ranged .70-.88	Factor analysis confirmed 4 dimensions of satisfaction (conducted with the parent version of the outpatient Satisfaction Scale); fit index = .974 Content validity - developed based on CASSP principles; Content area corrs w/ agreement re: termination from nonsignificant to .31				5-18 years, 12-18 for self report
Youth Satisfaction Questionnaire (Stuentzner-Gibson, Koren, & DeChillo, 1994)	Assesses level of satisfaction with services; 3 general questions; 2 items evaluating satisfaction with amount of help received; section for rating speciific services	Youth	Tested with predominantly white males	Few mins;	Cross & McDonald Help items minimally correlated with general satisfaction; Cronbach's Alpha = .80 for 3 general items; no stability information	No validity information does not look at components of care				9-18 years old receiving services

Self-esteem

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Canadian Self-Esteem Inventory for Children (Battle, 1976)	Measures self-concept; 30 items; yes/no format.		Tested w/ African Adols		Test-retest = .80 to .87; 2 day test-retest = .82 to .84	".81 to .89" w/ unknown variables				
Coopersmith Self-Esteem Inventories (CSEI; Coopersmith, 1967; 1982;); school form	Measures general self-concept, social self, home-parents, and school; lie scale; school form = 58 items; school short form = 25 items; adult form = 25 items; "like me" or "unlike me" format.	Child; adol	Initial norm grp from central Connecticut. Has means and standard deviations for White, Black, and Mexican American, SES, and community type.	10-15 mins; Consulting Psychologists Press, USA, 1981	Alpha (total) = .86, .61 - .71 for subscales. Test-retest (6 months) = .52 to .60 Test-retest (3 years) = .70 Internal consistency = .87 to .92 Split-half = .90 Alternate form reliability = .71 to .80	Corr w/: Children's Self-Concept Scale (r = .63); Behavioral Academic Assessment Scale (r = .47); Children's Social Desirability Scale (r = .17 to .33) Construct confirmed by factor analysis (Gilberts, 1983) Predictive: .35 to .53 corr w/ rdg achievement Corr w/ school achievement = .33; intelligence = .30	Normed on 1748 children	Test retest for ages 12-15 = .64; for 9-12 = .42; norms only for grades 4-8		Ages 8-15; adult form, ages 16+
Culture-Free Self-Esteem Inventories II (Battle, 1992)	Measures self-esteem: general, social, academic, and parents/home; lie scale; 60 items; yes/no format.	Self; individ or grp	Corr w/ parent's education = .16 to .17 (NS). Test developed w/ predominantly White, middle-class, Canadian children (Holaday, et al., 1996). Sig diffs btwn cultural groups: White, Catholic private school students sig more honest responses; sig diff scores from Hispanics and White students from Venezuela.	4th grade rdg Level; 15-20 mins; requires clinician w/ some training in administration and scoring of psychological tests; Intro kit = $124; $21/50 sheets There are 2 forms for grades 3-9; Holaday et al. - easy to score	Internal consistency child form = .66 to .76 Adult = .57 to .78 Alpha child = .76 to .92 Test-retest (child) = .81; adult = .82. Alternate form reliability = .70.	Corr w/ CSEI = .71 to .80 Corr w/ Adjective Generation Technique = .13 to .53 across diff samples Corr w/ perception of ability = .70 Corr w/ teachers ratings = .35 Corr w/ depression = -.55 Corr w/ nonverbal IQ = .08	Standardized on nearly 5,000 individuals	There are separate adult and child forms (grades 3-9); adols offended by some of childish wording - Holaday, et al., 1996 recc adol. form		Ages 5-65
Global Self-Worth Scale (subscale of the Perceived Self-Competence Scale for Children, Harter, 1982)	Assesses children's sense of self-worth; 6 items; 4-pt. scale.		Tested w/ Hispanic children		Alpha = .71 to .82					
Lipsitt Self-Concept Scale (Lipsitt, 1958)	Measures evaluations of self; 22 items; 5-pt.scale.	Child			Internal reliability = .84 to .85 in 3 administrations Test-retest = .54 to .91. Unknown reliability in one study = .81. Item-total corrs: 2 items below .20; 4 items below .30.	Corr w/ Children's Manifest Anxiety = -.34 to -.63 and positive relationship to academic motivation Lack of research on internal structure and validity (Francis, et. al, 1995) Corr = .25 w/ lie scale (social desirability responses) Corr = -.28 w/ neuroticism				9-21

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Multidimensional Self Concept Scale (MSCS; Bracken & Howell, 1991)	Assesses global self-concept in 6 domains of psychosocial functioning; 150 items or 25 items per domain.	Child (indiv or grp)	Results do not discriminate by ethnicity or gender	Approx 20 mins. Instructs clinicians how to use the results. Intro kit = $84; $44/50 forms. Requires M.S. to purchase; Pro-Ed, Austin, TX, USA.	Alpha >.90 for all subscales; >.97 for entire scale.	Face valid. Correlates "strongly" w/ other measures of self-concept and self-esteem; identifies clients previously identified as being low in self-concept.	Normed on a national sample of 2,501 children and adols btwn ages 9-19	Results do not desciminate by age		Ages 9-19 or grades 5-12
Offer Self-Image Questionnaire for Adolescents -Revised (OSIQ-R; Offer, Ostrov, Howard, 1982)	Measures adjustment and self-image; 129 items.	Self-report		**30-40 mins**						Ages 13-19
Perceived Competence Scale for Children (Harter, 1982)	Measures self-concept; 36 items.			**Older version**	Test-retest (3 months) = .70 to .87 for 208 students; .69 to .80 (9 months) for 810 students; Alpha = .75 to .80					
Perceived Competence Scale for Children (PCSC) (Harter, 1982)	Measures self-perception of cognitive, social, and physical competence and general self-esteem; 28 items; 4-pt. scale.	Child	Used w/ Chinese children in one study, and Ireland, and Australia	**Older version**	Alphas = .54 to .84			there are child and adol versions		
Perceived Self-Worth Questionnaire (author unknown)	Assesses perceived self-worth; 7 items; 4-pt. scale.				Alpha = .70	Corrs w/ other vars = -.03 to .25				
Personal Attribute Inventory for Children (PAIC; Parish & Taylor, 1978); revised by Parish and Rankin (1982) and renamed the Nonsexist Personal Attribute Inventory for Children (NPAIC)	Measures perceptions of self; 48 items; child selects 15 items which best describe self.	Child	PAIC study with normative data revealed gender differences in self-attributions, but norms are reported for males and females; these normative data are not applicable to the NPAIC. Test-retest coefficients differed by gender.		Test-retest (1 month) = .62 (NPAIC) across entire sample, but differed by grade and gender (some coefficients below .60). Split-half = .83.	Corrs w/ Behavior Rating Profile Student Scales (.26 to .48); State-Trait Anxiety Inventory for Children (-.44 to -.56); .66-.67 w/ Piers-Harris measure of self-esteem Child report - no signif diff btwn normals, phys handicapped, LD, and emot disturbed; but teacher report with significant differences	Normative sample of 1,050 children with separate norms by gender; however, the nonsexist version was introduced after the normative study; also a 1,400 college student norm sample			

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Piers-Harris Self-Concept Scale (Piers, 1969; revised manual, 1984). Revised version available. Also called: Children's Self-Concept Scale; and The Way I Feel About Myself	Measures self-esteem and problem areas in self-concept, 6 subscales, 80 items, true/false or yes/no format.	Child or adol	Has Spanish version. Hammil, Brown, and Bryant report inadequate demographics for normative sample; the manual itself states that the population is limited. Has been used w/ Hispanic and Black children. Samples of Indian children in Florida and Montana had sig lower scores than the normative White sample. Some samples of Blacks and Asians scored well above the mean. Samples of majority and minority ethnicity w/in the same study often w/ similar scores, suggesting that scores vary w/ environmental context. Corr w/ academic grade average = .43 for Indian children but only .16 for White children. Lower test-retest for minorities.	15-20 mins; 3rd grade rdg level; kit=$99.50; $29.50/20; Western Psychological Services, Los Angeles, CA, USA	Internal consistency = .78 to .93; subscales = .73 to .81; Internal consistency w/ Kuder Richardson = .95 to .97; .71 to .86 for subscales test-retest = .65 to .81for public school populations; .42 to .96 for handicapped and ethnic miniorities ten week test-retest = .75	Content validity; items w/ low discriminatory power were dropped (Long & Hamlin, 1988); items generated from children's statements of what they liked and disliked about themselves. Construct confirmed by factor analysis (Gilberts, 1983). 10 factors accounted for 43 to 45% of variance. Evidence of factorial instability across diff populations and even in the same population at 10-week retest (Platten & Williams, 1981) Corr w/ Lipsitt CSCS = .68; .41 w/ teacher ratings; .49 w/ peer ratings. Corr w/ Body Esteem Scale = .68. Corr w/ parental love = .56; peer acceptance = .61. Corr w/ attitude toward school = .22. Corr w/ academic grade average = .43 for Indian children but only .16 for White children. Corr w/ anxiety = -.54 to -.69; health probs = -.48; big problems = -.64. Corr w/ Behavior Problem Checklist = -.34. Predictive corr w/ IQ = .25; achievement = .32. Divergent: corr w/ nonverbal intelligence = .08.	1700 students from 7-18 for norms; but Hammil, Brown, & Bryant report norms before 1976; in the early 1960s, 1183 school children grades 4-12	Age-stratified norms		Ages 8-18

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Rosenberg Self-Esteem Scale (Rosenberg, 1965)	Measures self-acceptance; 10 items (4 point scale).	Adol		3-5 minutes; Princeton University Press, USA, 1965	Test-retest = .62. Alpha = .73 to .83 for grades 7-12 across studies; .84 for middle school children.	"Negative relationship" w/ depression, neurosis, and psychosomatic symptoms. "Positive relationship" w/ student leadership, reflected self, and ability to criticize self. In a multi-trait, multi-method study with 11th and 12th graders, the scale corr .64 with a semantic differential method of measuring general self-concept, .79 with a Likert scale (SDQ-III) Discriminant validity off-diagonal values ranged .19 to .54. Harter's SPPC Global Self-Worth corr = .72, .66, and .66 w/ Rosenberg total, pos self-esteem, and neg self-esteem, respectively. Corr w/ self-esteem items = .06 to .41 Did not signif differentiate high and low track students.	5,024 high school students in 1965; as of 1983, no new norms available	No diffs btwn grades 5-8 in scores		Grades 5-8; adol or adult
Self-Appraisal Inventory (Narikawa & Frith, 1972)	Measures school child's self-concept: scholastic, peer, family, and general; 36 items; yes/no format. Also has 27 and 64 item versions that have been used	Read to child by an examiner and completed by child	Children from ethnically mixed families displayed lower self-concepts		Test-retest (2 weeks) = .29 to .58 across subscales, .73 on the total score. KR-20 coefficients = .50 to .62 across subscales, .37 for total scale. Alpha = .41 to .65 across subscales, .82 on the total score.	Corr w/ CDI and RCMAS >.60 Factor analyzed Signif differentiated referred and nonreferred children Corr w/ Classroom Adjustment Rating Scale subscales = .11-.37 Corr w/ CP (sociometric measure) subscales = .04-.70				Children
Self-Description Inventory (Mboya, 1993)	Measures self-concept; 50 items; 5-pt. scale.			**40 minutes**	Alpha = .74 to .90 across subscales.	Corrs btwn subscales from .02 to .32; 8 factors w/ eigenvalues >1.00; squared multiple corrs of 8 factors = .70				Adols

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Self-Description Questionnaire I (SDQ-I; Marsh Parker, & Smith, 1983)	Measures academic, non-academic, and global self-concept; 72 items (56 positively worded items, 10 negatively worded items, and 6 experimental general self-concept items that have been dropped occasionally by author); 5-pt. true/false scale.	Child	Stronger relationship w/ academic ability and student-teacher agreement among students with higher SES and in schools in which students had higher academic ability.	15-20 mins; Psychological Corporation, USA, 1987	Alpha = .65 to .95 across subscales and grade levels.	10 negatively worded items created own factors w/ Alpha = .73 (unwanted bias according to author); also .27 corr of negative worded items to positive items. Factor analyses from several studies have identified the same 7 factors; factor analysis of 56 positive items hold same 7 factors. Math self-concept corr w/ math achievement = .55. Rdg self-concept = .21 corr w/ math achievement. All school subjects self-concept = .43 corr w/ rdg; other self-concepts not related to rdg		Negative worded items very difficult for youngest children = -.02 corr w/ positive worded items; .60 corr for older children; From grade 2 to grade 5, several subscales found to vary w/ age.		Ages 6-12
Self-Description Questionnaire II (Marsh, Smith, & Barnes, 1983)	Measures academic, non-academic, and global self-concept; 56 items; multiple choice and rating scales. 102 items according to a 1997 article (see Marsh, 1992)	Adol		Approx 20 mins; Psychological Corporation, USA, 1990	Alpha = .80 to .93 across subscales	corr w/ academic ability = .3 to .7; corr w/ success/ability = .59; success/effort = .55; failure/ability = -.48; failure/effort = -.40 Corr rdg achievement w/ rdg self-concept = .31; math self-concept not corr w/ rdg achievement Confirmatory factor analysis did not fit data well - Taylor Lewis Index <.90 (.826 to .883)		Alphas calculated by age = .84 to .90 grades 6 to 10		Grades 7-10
Self-Description Questionnaire III (SDQ-III; Marsh et al., 1990)	Measures 4 areas of academic self-concept, 8 areas of non-academic self-concept, and a general subscale; 136 items.	Older adol or young adult	Hammil, Brown, & Bryant report inadequate demographics for normative sample	20-30 mins; Australia, 1989	Internal consistency consistently in the 0.80s and 0.90s across 3 administrations, w/ the exception of the honesty subscale (.72 to .77); test-retest (1-month intervals) for Time 1/Time 2, Time 1/Time 3, and Time 2/Time 3 = .77 to .94, .66 to .92, and .73 to .94, respectively	Hammil, Brown, & Bryant report that the SDQ-III is not validated	Normed on over 1000 students w/ at least 100 per age		"Responses across all 13 SDQ III factors differed substantiallly w/ the time of testing" (Marsh, Richards, & Barnes, 1986)	Grades 7-12

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Self-Esteem Index (SEI; Brown, & Alexander, 1991)	Measures the way youth perceive and value themselves: academic competence, family acceptance, peer popularity, and personal security subscales; 80 items; 4-pt. scale.	Child or adol	Hammil, Brown, & Bryant report good demographics for normative sample	Approx 30 mins; requires clinician w/ some training to administer and score; intro kit = $109; $68/50 answer and scoring forms; PRO-ED, USA, 1991	Test-retest > .80. Internal consistency >.80 for all subscales.	Hammil, Brown, & Bryant report at least 5 types of empirical validation.	Norm referenced on over 1000 subjects w/ at least 100 per age	Alphas approach or exceed accepted standards for one year intervals		Ages 7-18 years, 11 months
Self-Perceived Role Competence (Beiser, Lancee, Gotowiec, Sack, & Redshirt, 1993)	Measures self-perceived instrumental and social competence.		Used w/ four native populations in US and Canada; ethnicity by item interaction effects small: 0.8% for social competence; 0.008 for instrumental competence		Internal consistency ranges: social competence (native children, ∀ = .61-71; non-native children, ∀ = .60 to .78); instrumental competence (native children, ∀ = .58 to 76; non-native children, ∀ = .78 to .87); test-retest (1 year) = .33 to .58 (by grade)	Convergent validity w/ teacher performance ratings = .21 to .35 (native children) and 35 to .58 (non-native children); convergent validity w/ rdg, writing, and math achievement = .12 to .41 (native children) and .27 to .65; corrs (.41 and .61) w/ global self-concept (Harter, 1982)				
Self-Perception Profile for Adolescents (Harter, 1988)	Assesses self-perceptions w/ 9 subscales: scholastic competence, social acceptance, athletic competence, physical appearance, job competence, romantic appeal, behavioral conduct, close friendships, and global self-worth; 45 items; 4-pt. scale.		Has Norwegian version; Factor structure confirmed in Australia except for 2 factors which merged;	University of Denver, USA, 1985	Alpha = .56 to .87; Job subscale w/ internal consistency of .41, romantic only .65 and conduct only .68 after one item was dropped, other 6 scales = .72 to .90	Corr social acceptance scale w/ peer sociogram = .42; w/ friendship = .22; w/ loneliness = -.94; w/ grades = .01 Corr close friends scale w/ friendship = .33; w/ loneliness = -.65; w/ grades = 0 Corr self-worth scale w/ psychological probs = -.52; w/ quality of life = .52 Corr athletic competence scale w/ training = .42; w/ grades = .08; Corr physical appearance scale w/ body satisfaction = .84; w/ grades = -.02; Corr romantic appeal scale w/ boyfriend/girlfriend = .39; w/ sexual experience = .50; w/ grades = -.06 Corr Scholastic competence scale w/ grades = .54; w/ friendship = .06 All scales corr w/ Behavioral adjustment = .07 to -.43 factorial validity - 7 factors >1.0; 6 kept - explained 55% of variance Corr w/ social desirability = -.05 to .26 across subscales				Ages 13-19

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Self-Perception Profile for Children (SPPC; Harter, 1985) **Revision and extension of the Perceived Competence Scale for Children**	Taps 6 domains of self-esteem: global self-worth, scholastic competence, social acceptance, athletic competence, physical appearance, and behavioral conduct; 36 items; 4-pt. scale.	Child; **perhaps parents and teachers can also report on child (Levendosky et al. did it)**	Used in Dutch child psychiatric clinic; Dutch version replicated factor structure and reliability; Dutch norms available (361 children); but separate Dutch study, goodness of fit = .93 to .96 but some items loaded on illogical factors Used in Taiwanese urban and rural schools; used w/ Italian children; Scottish children, reliability and validity data available for German version; used w/ protestant and catholic Irish adols - held same factor structure for protestants and catholics; used w/ Canadian children - factor structure held again except for 2 items; used in Chinese (translated w/ back translation checks) - factor structure did not hold - 12 items migrated to other factors; also sig lower scores than Canadian children for 4 of 6 subscales	University of Denver, USA, 1985	Internal consistency = .71 to .86 for the subscales over several studies; test-retest (4 weeks) for Dutch school sample = .66 to .83; Cronbach's Alphas for Dutch school sample = .65 to .81; test-retest (1 year) = .28 to .61 2 year test-retest on subscales = .27 to .63 2 year corr w/ SPPA for children going to adols = .28 to .63, .21 for conduct subscale which changes on the adol measure	5 factors w/ eigenvalues >1, minimal cross loadings across factors, and account for 48 - 55 % of variance; 6 items per scale loaded on approp factor >.40 - one exception w/ one item in one study & 2 items in another study; factor structure consistent for clinical and nonclinical group Corr .31 w/ percent of competent solutions generated in social problem situations; parent ratings of social competence subscale corr .32 w/ total number of solutions generated, -.23 w/ punitive responses, and .24 w/ adapative responses; teacher ratings of social competence subscale corr -.36 w/ punitive responses Harter's SPPC Global self-worth corr .72 w/ Rosenberg total , .66 w/ Ros pos self esteem and, .66 w/ Ros neg self esteem corr perceived scholastic competence scale w/ math achievement = .17 to .35	Normative data			Ages 8-13
Tennessee Self-Concept Scale (TSCS; Roid & Fitts, 1988)	Measures self-concept w/ regard to identity, behavior, self-satisfaction, family relationships, and social relations; 100 items; 5 pt. scale; self-descriptive statements that the adols use to portray their concept of self	Adol	Normative sample unrepresentative College-level students heavily represented in normative sample; and mostly Whites	10-20 mins; TSCS II offers guidance to clinician for interpretation and treatment; 3rd grade rdg level, 2nd grade level for child form; $130/kit; $32.50/20 forms; Western Psychological Services, USA, 1991; software available; considered complex to score and needs computer scoring (Gilberts, 1983)	Version II psychometrically equivalent to I???; TSCS I - test-retest = .80 to .92 Alpha = .92 to .94	Little support for scale's ability to measure internal dimensions of self-concept as originally hypothesized (Buri, et. al, 1992) TSCS I: prediction of anxiety = -.70; medical probs = -.56; feelings = .64; found to discriminate patients and nonpatients, delinquents and nondelinquents, 1st time and repeat juvenile offenders, alcoholics and nonalcoholics	Original norms - sample of 626 from around the country II has norms down to age 7 - sample of 3000	Child (7-14) and adult (13 and up) forms		Ages 12 and older; TSCS-II ages 7-90
What I Think About Myself (Mathewson, Orton, 1963)	Measures self-esteem; 18 items; agree-disagree format.			10 mins; semi-projective instrument	Test-retest = .75 to .96 across grade levels.			Test-retest = .75 to .96 across grade levels		Ages 5-23

Note: Pos = positive; Neg = negative; corr(s) = corr(s); grp = group; indiv = individual; w/ = with; rdg = reading; SES = socio-economic status; sig = significant(ly); diff(s) = difference(s); btwn = between; var(s) = variable(s).

Social Support

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Family, Friends, and Self Assessment Scales (Simpson & McBride, 1992)	Measures peer involvement, involvement with peers who get into trouble, and parent familiarity with peers	Youth								
Perceived Social Support Scale from Family (PSS-FA) and Perceived Social Support Scale from Friends (PSS-FR; Procidano & Heller, 1983); Perceived Social Support from School Personnel (DuBois, Felner, Brand, Adan, & Evans, 1992)	Measures perceived social support received from family and friends; a social support received from school personnel version was developed for one study; each scale has 20 items; yes/no format	Adolescents			Alpha = .85 to .90 for all Corr PSS-FA w/ PSS-FR = .23	Factor analysis of PSS-FA and PSS-FR support the uni-dimensionality of each scale. Corr w/ psych distress = -.15 to -.63; all 3 scales corr w/ Self Appraisal Inventory = .31 to .44; all scales w/ CMAS and CDI = -.16 to -.45; corrs w/ GPA: fam = .15, friend = .26, school = .06; PSS-FA corr w/ identity = .69; .60 w/ self reliance; .75 w/ work orientation; PSS-FR corr w/ identity = .37, .26 w/ self reliance, and .17 w/ work orientation.				adolescents
Personal History Inventory for Children (Parish & Wigle, 1985)	Measures parental support systems, 14 items, yes-no responses, scored by summing “yes” responses	Child, parent, or teacher self-report on the functionality of the child’s family support system			Test-retest (1 month) = .80 (divorced parents) and .93 (intact families); 1 month test-retest (Parish, 1987) = .55 (with college students)	Correlates significantly with self-evaluations, mothers’ evaluations, & fathers’ evaluations (.18 - .41)				
Social Support Appraisals Scale - Revised (Dubow & Ullman, 1991)	Asesses peer, family, and teacher support; 41 items (5 pt. scale)	Child	Corr w/ race = .18; -.12 w/ SES		Alpha = .86 to .93;	Corr w/ teacher or parent rated probs = -.19 to -.28 Corr w/ teacher rated competence = .19 Corr.w/ original version from 2 years earlier = .38 Same 3 factor solution as original scale w/ alphas = .86 to .92; New items corr w/ old version items = .84 2 years later				
Social Support Appraisals scale (Dubow & Ullman, 1991)	Assesses peer, family, and teacher support; 31 items (5 pt. Scale)	Child	.00 corr. w/ race, -.16 corr w/ SES	**Older version**	Alpha = .78 to .88; 3-4 week test-retest = .75	Corrs with parent or teacher probs = -.22to-.30; Corr w/ teacher rated competence = .38 3 factor solution w/ alpha’s .78 to .83				8 to 18

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Social Support Scale for Children and Adolescents (Harter, 1985)	Measures parent support, teacher support, classmate support, and close friend support; 24 items (4 pt. scale)			Close friend support has been dropped in studies because many kids do not have a close friend	Internal consistency = .72 to .88	Corr CBCL total w/ parent support = .06; -.28 for teacher support; .30 for classmate support Harter child report behavioral conduct scale corr w/ parent support = .82; .67 w/ teacher support; .01 w/ classmate support; Across studies: Global self worth corr w/ parent support = .43 to .50; .35 to .64 w/ teacher support; .35- .48 w/ class support;.38 w/ close friend Corr Social acceptance w/ parent support = .08; .46 w/ teacher support; .53 w/ class support Corr w/ Social Support Appraisals Scale = .56 to .66 across subscales				
Supportive Parenting Scale (Simons, Lorenz, Conger, & Wu, 1992)	Measures parental support; 9 items (5 pt. scale)	Child self-report, mother and father report			Alpha = .78 to .89 Corr. mother's supportive parenting w/ father's = .62-.78	Corr. w/ harsh discipline = -.34 to -.51				

Stressful Events

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Academic Pressure Scale for Adolescents (West & Wood, 1970)	"Identifies the perceived antecedents of academic stress" (Jones, 1993); 35 items, 5 pt. Likert scale				Test-retest = .78		Girls reported significantly higher levels of stress on 8 of 35 items			Adolescents
Adolescent Perceived Events Scale (Compas, Davis, Forsythe, & Wagner, 1987)	Assesses occurrence and impact of stressful events; different versions for early, middle, and late adolescents (159, 200, and 210 items, respectively); respondent indicates whether an event occurred in the last 3 months; if an event occurred, young adolescents rate the desirability of those events; middle and older adolescents rate desirability, impact, and frequency of events; 9 pt. scale for rating events that occurred	Adolescents	Homogeneous sample with regard to ethnicity and living environment	Administration time unknown, but probably lengthy; Compas and associates (1987) assessed the impact of length (potential fatigue) on reliability scores	Test-retest (2 weeks) for occurrence of stressful events (across 3 age-groups) = .77 to .85 for total number of events, .74 to .89 for weighted negative events, and .78 to .84 for weighted positive events; percent agreement of the occurrence, desirability, impact, and frequency of stressful events ranged 83 to 89%, 81 to 93%, 77 to 90%, and 90 to 91%, respectively; reliability was unaffected by length/ fatigue. In an older adolescent sample, cross-informant agreement (percent agreement with college roommate as rater) was 82% for the occurrence of stressful events, 87% for ratings of desirability, 90% for impact, and 91% for frequency.					Early adolescents (ages 12-14), middle adolescents (ages 15-17), and late adolescents (ages 18-20)
Children's Own Perceptions and Experiences of Stressors (Colton, 1989)	Measures children's perceptions, experiences, and emotional reactions to stressors, 60 items, 5-pt. scale	Child			Average alpha for 7 factors = .84					

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Daily Hassles Questionnaire (Rowlison & Felner, 1988)	Assesses daily life events that have been a hassle over the last 2 weeks; 81 items; 4-pt. scale		During item generation, included low income subjects in sample		Alpha = .95.	Content validity: 44 items rated by Masters and Doctoral Clinical Psychologists as overlapping with signs of emotional distress Developed after obtaining open-ended responses. Corr w/ SAI, CMAS, and CDI = -.45, .48, and .50 respectively. Corr w/ Life Events Checklist = .28 to .33; w/ psych distress = .33 to .58. Corr w/ negative life events = .44. Corr w/ family support = - .28 to -.38. Discriminant validity: corr w/ GPA = -.02 to -.24.				School age children and Adolescents
Family Inventory of Life Events (McCubbin, & Patterson, 1983)	Assesses the pile-up of normative and non-normative life events and changes experienced by a family in the past year; 9 scales and an overall stress index; 71 items	Parent or adol self-report			Alphas across 3 large samples = .79 to .82 (full scale), .71 to .73 (intrafamily strains), .09 to .20 (marital strains), .22 to .27 (pregnancy and childbearing), .58 to .61 (finance and business), .55 to .56 (work-family transitions), .51 to .60 (illness and family care), .31 to .60 (losses), .48 to .55 (transitions), and .44 to .71 (legal). Parental agreement = .42; parent-adolescent agreement = .29 to .35	Convergent validity: 10 subscales correlated with the Family Environment Scale				Families with dependents; adolescents
Hassles Scale for Children (Parfenoff & Jose, 1989)	Determines type and degree of hassles in children's daily lives; 49 items; 3-pt. scale	Child			Test-retest (2 weeks) = .74. Alpha = .88.					

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Life Events Checklist (Johnson & McCutcheon, 1980)	Measures stressful life events; 46 items; 4-pt. scale.					Corr w/ Daily Hassles Questionnaire = .28 to .33; w/ psych distress = .27 to .35. Negative life events corr w/ DHQ= .44. Good events corr w/ Self Appraisal Inventory = .08; -.01 to -.03 w/ CMAS and CDI; bad events corr w/ self appraisal = -.28; .29 to .30 w/ CMAS and CDI. Discriminant validity: corr w/ GPA = -.09 to -.21; good events corr w/ GPA = .10; bad events corr w/ GPA = -.16.				Older children and adolescents
Life Events Scale for Children (Coddington, 1972)	Determines quantity of major life stressors child has experienced in past 12 months; 30 items	Parent report about child	Spanish version based on Coddington (1972)	15 mins.	No additional information found.	Little info on reliability (Carson et al., 1992).				Pre-school age children; perhaps older children in another version
Sandler and Block's (1979) Modification of the Coddington Life Events Scale for Children	Life events within past year of child's life; 32 items	Parent	.16 corr w/ race; .25 corr w/ SES		2 year test-retest = .49	Content validity: items derived from literature; items judged needing readjustment for elementary school children by teachers, pediatricians, and mental health professionals 21 to .36 corr w/ teacher and parent problem ratings.	Has norms			

Therapeutic process

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Adolescent Working Alliance Inventory (AWAI; Linscott, DiGiuseppe, & Jilton, 1993)	Measures therapeutic bond, agreement on therapy goals and tasks; authors suggest using 36 items until research warrants reducing the scale; therapist version unchanged from original WAI (36 items)	Adolescent or therapist		Readability ensured by revising items until understood	Internal consistency with 23 total items: Alpha for subscale items ranged from .90-92 (adolescent and therapist versions); Alpha (total) = .96 (adolescent and therapist); cross-informant correlations ranged .38-.48 for subscales, .49 for total	Content validity: revised from the adult version; dropped 13 of original 36 items that correlated < .40 with other items in the subscale. Factor analysis (adolescent): 3 hypothesized factors accounted for 68% of the variance, but only 10% for the last 2 factors; "therapeutic alliance for adolescents may be one [sic] dimensional" (Linscott et al., p. 8); (therapist) 3 hypothesized factors accounted for 68% of variance, but only 12% for the last 2 factors (again, suggesting uni-dimensionality). Adolescent version correlated -.19 to -.22 with session number; therapist version correlated -.27 to -.33 with session number; correlations (both versions) with depression ranged .05-.12; corrs with YSR int ranged -.13 to -.26 (adolescent) and .10-.20 (therapist); corrs with YSR ext ranged .00 to -.10 (adolescent) and .18-.31 (therapist).				
Barriers-to-Treatment Participation Scale (Kazdin, Holland, & Crowley, Breton, 1997)	Assesses barriers that can occur over the course of treatment, stressors and obstacles that compete w/ tx, treatment demands, perceived relevance of tx, therapeutic alliance, events leading to tx termination; 58 items	Interview with parent in person or by telephone; also therapist version			Corr parent w/ therapist = .39 to .45; Spearman-Brown coefficient across informants = .86 to .89 Alpha = .86 to .93 for both informants on total score across studies; Alpha for subscales = .61 to .87 for both informants across studies Item-total corr = .11 to .55 for parent and .06 to .75 for therapist version	Content validity - items developed thru focus group discussions with therapists Significantly discriminated dropouts from completers (in 2 studies); effect sizes by subscale for parent and therapist measures: stressors =.70-.72 & 1.25-1.26; tx demands = .11-.12 & .58-.65; relevance of treatment = 1.07-1.10 & 1.67-1.70; relationship w/ therapist = .39-.43 & .55-.61 Also significantly discriminated by length of treatment, familes that cancel sessions, and no-shows; critical events scale did not significantly discriminate groups Factor analysis - 15 parent factors and 11 therapist factors w/ eigen value >1; 66-69% of variance; scree test found large drop in eigen value after 1st factor and all other eigen values at asymptote - thus used total score w/o separate factor scores				3-13; children and fams referred for outpatient tx
Family Involvement items (Baker, Blacher, & Pfeiffer, 1993)	Measures four dimensions of family involvement plus an overall index (sum total): participation in social and educational activities at facility (3 pt. scale), frequency of phone calls to staff or child, frequency of visits to child, and frequency of visits home or overnight with family (last 3 domains on a 6 pt. scale); 4 items	Residential facility staff member who knows child rates family	Family involvement index corr -.50 with SES, .34 with race, and .22 with parents' marital status	Items embedded in client data forms with 13 additional items	Alpha for family involvement index = .87	Family involvement rated lower among families with children with dual diagnoses (3 of 4 domains and overall index). Family involvement index corrs with distance from facility = -.57 and state (family residence) -.50.				Any child or adolescent 19 years of age or younger

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Loyola Child Psychotherapy Process Scales (Estrada & Russell, 1994)	Assesses positive and negative aspects of child and therapist behaviors- ratings of 15 min. segments of therapy; 33 items - 18 child therapist items, 15 child items (5 pt. scale): overall rating is positive score minus negative score	Independent raters of child and therapist items			Interrater reliability = .72 to .85 for subscales across child and therapist; 90 % of paired ratings within one scale point	Factor analysis accounted for 67-73% of variance across child and therapist Several factors significantly differentiate good and poor therapeutic sessions				
Motivation to Change Questionnaire (Phares & Danforth, 1994)	Measures motivation to change behaviors endorsed on the Youth Self Report (YSR; Achenbach & Edelbrock, 1987); yes-no format	Adolescent self-report				Corrs with ratings of how bothersome is the behavior: .60 (internalizing) and .78 (externalizing)				
Penn Helping Alliance Questionnaire (Alexander & Luborsky, 1986)	Measures client's perception of therapist as supportive and helpful and perception of the relationship as a working collaboration (therapist version measures the therapist's perspective of the quality of therapeutic relationship); 11 items; 7 pt scale ranging from -3 to +3	Adolescent and therapist self-report			Alpha = .92 with inpatient adolescents	"...one of the most consistent predictors of treatment outcome across types of patients and types of treatments" (Eltz, Shirk, & Saiz, 1995, p. 421); signif differentiated maltreated from non-maltreated adolescents alliance formation corr. w/ interpersonal problems = -.39 to -.46 across reporters high and low alliance formation significantly differentiated on therapist rating of outcome				
Psychotherapy Process Inventory (modified for use with children) (Gorin, 1993)	Therapists described use of specific techniques used throughout therapy; 5 pt. frequency of use scale; 16 client involvement/ motivation items	Therapist		Training required but amount of time not reported, interviewers were closely supervised and retrained every 20 sessions???	Alpha = .92 for client participation items; interrater r=.76; Alpha = .83 for therapist directive support items; interrater r = .88 Interrater for entire scale - kappa >.70 (w/ extensive training and supervision)	Content validity - adult version w/ factorial validity, and reliability (Gorin, 1993) Client participation scale corr w/ Global impairment change = .45				
Smith-Acuna, Durlak, & Kaspar (1991) unnamed scales	Measures 4 dimensions of the therapy process: child affect, therapist behavior, perception of the other's affect, and goals for session; data collected via Q-sort and open ended questions; 69 items for therapist; 39 items for child	Trained examiner administers to child and therapist fills out independently		10 to 15 min.	Internal consistency therapist - .61 to .88 across scales; Child - .68 to .86 across scales Child and therapist reports not significantly related (no data provided)	Content validity - adapted from adult instruments; items correlated <.20 w/ total scale dropped; items derived from an adult form Factor analyses were not performed Subscale section - goals for child report provided unusable data (authors)				6 to 12

Measure	Description	Informant	Cultural relevance	Suitability	Reliability	Validity	Norms	Developmental sensitivity	Sensitivity to change	Target population
Therapeutic Alliance Scales for Children (Shirk & Saiz, 1992);	Measures 3 elements of therapeutic alliance: child's affect toward therapist (bond and negativity) and 1 item for collaboration on tasks ("talking about problems" or "verbal"); therapist and child versions; 9 total items	Child and therapist			Cronbach's alpha = .67-.74 (child), .72-.88 (therapist); agreement (corr) on bond, negativity, and verbal subscales = .42, .37, and -.12, respectively	Content validity: items based on theoretical and empirical grounds; items solicited from experienced child clinical psychologists and psychiatrists; items w/ low item-total corr. eliminated absolute value corrs (child and therapist) with therapist global rating of child's participation = .03-.11 and .28-.57, respectively construct valid - bond & negativity supposed to be neg. related; corrs = -.50 to -.57; across informants = -.19 to -.27;				Used with ages 7-12
Therapy Process Checklist (Weersing, 1996)	Measures "therapist use of different treatment techniques..." and "etiologic view of psychopathology" in 3 different psychtherapeutic orientations; 85 items	Therapist self-report			Alpha = .91-.97 for technique scale and .69-.94 for etiology scale (across theoretical orientation)	Content validity: items reflect 3 therapist-described theoretical orientations; items generated from review of empirical and theoretical articles; items generated when a behavior appeared in at least 4 articles. Factor analysis yielded hypothesized 3-factor structure for both scales. Partial correlations with self-reported orientation for technique and etiology scale = .63-.86 and .39-.78, respectively; however, only the psychodynamic orientation finding replicated (significantly) in a clinical sample (actual data not provided)				Child mental health providers
Treatment Acceptability Questionnaire (TAQ; Hunsley, 1992)	Assesses the treatment acceptability of psychological treatment and can be used across therapeutic modalities; 6 items; 7 pt. scale	Child self-report			Used a convenience sample of psychology undergraduate students to rate a college student case vignette and another sample to rate a child case vignette; 3-week test-retest = .78 (college student case vignette). Internal consistency was .74 and .80 across administrations (college student case vignette), and .81 (child case vignette).	Content validity: items based on literature. Correlated .87 with TEI (Kazdin, 1980), .59 with Evaluative (good-bad) subscale of the Semantic Differential method. Correlated -.02 and -.06 with Potency (strong-weak) and Activity (active-passive) subscales of the Semantic Differential method (not hypothesized as divergent constructs). Discriminated between a cognitive-behaviorally based intervention and symptom prescription approaches using a case vignette.				
Treatment Evaluation Inventory (TEI; Kazdin, 1980) and TEI-Short Form (TEI-SF; Kelley, Heffer, Gresham, & Elliott, 1989)	Rates child's treatment as acceptable, likable, suitable, cruel, unfair, and effective according to parents; 15-item version (7-pt. scale); 9-item simplified version (5-pt. scale).	Parent	Signif diffs by parent income and race. Mothers and fathers rated the acceptability of 5 of 6 child management techniques differently.	TEI: 3-4 mins; TEI-SF: 2 mins. 71% of subjects liked TEI-SF better than TEI. TEI: 5th grade reading level; TEI-SF: 4th grade reading level.	Median inter-item correlation = .67. TEI alpha = .89 TEI-short alpha = .85	Content: factor analysis used to design measure. Factorial: High factor loadings; separate factor analyses found 2 factors instead of original one - 42 and 15-19% of variance per factor; congruence of factors across studies for each factor .95 and .87; on TEI-SF, items loading >.30 on both factors were dropped; 2 factors explained 57 and 12% of variance (69% total). Short and long versions yield unequal evaluation scores for some types of treatment. Correlated .87 w/ TAQ Significantly differentiated different child treatments across studies for long and short forms; response cost and positive reinforcement rated superior to other child management techniques; discriminates between alternative treatment methods (Miller & Kelley, 1992).				

APPENDIX 6

HOW TO MEASURE CHANGE

Repeated Measurement: Needed to Evaluate Change

Warren Lambert & Len Bickman

Executive Summary

Almost all evaluations of mental health outcome rely on pre-post experimental or quasi-experimental designs. The present appendix presents new evidence about problems with pre-post analysis and shows how multi-wave repeated measurements, when analyzed with the appropriate longitudinal models, solve these problems.

Multiple wave data offers a) better statistical power; b) important information about when things occur; and c) solutions to the statistical paradoxes that necessarily occur when only two data points are used to answer three questions: a) Did the whole sample improve over time? b) Were the groups equal in severity at intake? and c) Did the experimental groups improve more or improve faster?

After demonstrating how difficult it is to predict a client's future endpoint, a repeated-measurement outcome monitoring method is presented. The method charts individual client progress over time and raises a warning flag if a client's outcome is worse than expected based on that child's progress and the progress of other children in the sample.

The proposed method of monitoring, we believe, may someday be a cost-effective addition to continuous quality improvement procedures designed to detect adverse outcomes and trigger corrective actions in mental health services in ordinary community settings.

Table of Contents

To Evaluate Outcome Use 3-6 Repeated Measures

Abstract. Adding additional waves of data collected from each participant increases power, especially when a two-wave design is increased to four or five waves. This improvement is a result of the fact that having an increased number of data points reduces the standard error of estimate. More precision means more statistical power to detect differences between groups.

Introduction

Despite decades of inconclusive debate about change scores, the pre-post quasi-experimental (Campbell & Stanley, 1963) design is all-too-alive and well in outcome evaluation. The design's face validity gives it undeniable appeal, especially for under-funded program evaluations studying effectiveness in field settings.

In the last decade, classical change scores such as differences and residuals have been increasingly replaced by multi-wave longitudinal models (Bryk & Raudenbush, 1992; Diggle, Liang, & Zeger, 1994; Gibbons, Hedeker, Elkin, Waternaux, et al., 1993). These new statistical models make the analysis of multi-wave data convenient and attractive.

The first study of this appendix uses Monte Carlo simulation to see whether multi-wave outcome analysis offers significantly better statistical power than classical methods based on the pre-post design and analysis of variance.

Method

The present study examined power to detect a difference between two experimental groups in a longitudinal study. Two questions were of special interest: a) How does the pre-post design compare with a longitudinal design with repeated measures; and b) how much power do the intermediate waves between pre-test and post-test contribute?

These questions were answered with Monte Carlo simulation data calculated so as to resemble the Vanderbilt Fort Bragg Evaluation Project (FBEP) data set as closely as possible.). The FBEP is a 5-year longitudinal study, intake plus 6 follow-ups. The observed 7-wave means and variance/covariance matrix for the CBCL Total Problem Score in the FBEP built into the simulation using the method of Khattree & Naik (1995, p. 21)

Results

The simulation produced means very close to real means of the FBEP, more often than not within 0.1 SDs of each other. The SDs are extremely similar, only one differing by 0.1 or more. Because the cross-wave correlations influence the power of longitudinal analyses

(Diggle et al., 1994), the simulated data was generated by a routine (Khattree & Naik, 1995) that could duplicate any given variance/covariance matrix. Simulated cases were generated to duplicate the variance-covariance matrix of the FBEP CBCL. The Monte Carlo data was similar to actual CBCL scores from a 5-year longitudinal study. This similarity would suggest that results could be applied to studies of similar structure. Power estimates for studies with different covariance structures may differ considerably.

A group difference in improvement over time was built into the data; the effect size was made large. The experimental group had a linear improvement advantage of two points per year or 10 points (1 standard deviation) at year five. A difference of 0.80 SDs between means is considered large (Cohen, 1988; Cohen, 1992).

The experimental designs studied include the two-group pre-post difference score and longitudinal designs with from 3 to 7 repeated measurements. The pre-post design was tested with repeated measures analysis of variance (ANOVA); the longitudinal designs analyzed with random coefficients longitudinal models (Rogosa & Saner, 1995). The null hypothesis stated that the two groups (experimental and control) improved equally over time. Power curves appear in Figure 1, which shows six power curves, pre-post ANOVA and five longitudinal models.

The difference score ANOVA was surprisingly not the least powerful. The black power curve was based on a pre-post change score using repeated measures analysis of variance (ANOVA). For this classical analysis, about 70 subjects are required for 80% power.

The longitudinal designs were more powerful than the ANOVA. With a full 7 waves in the analysis, 80% power is reached at less than 35 subjects, half the number required for 80% power compared with the pre-post ANOVA. There were differences among the longitudinal analyses as well. With only 3 observations per subject, about 50 subjects are needed for 80% power, compared with less than 35 for the 7-wave model. All of these curves in Figure 2 suggest a single result: More data provides more power, regardless of whether “more data” means more subjects or more observations per subject.

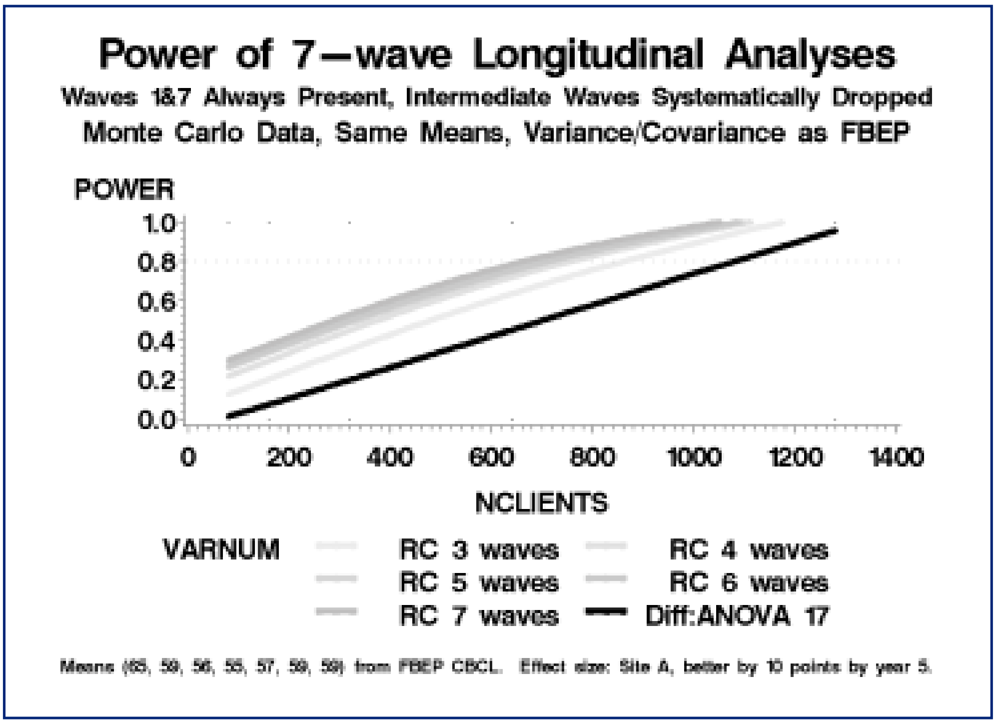


Figure 1.

When we add subjects to a study, the between-subjects standard error of the mean shrinks by **SEM = SD/sqrt(N)**, i.e. as N goes up the standard error become smaller. This increase in precision shrinks the 95% confidence intervals so smaller differences can be detected and power increases. To determine whether a similar effect occurs when extra observations are added for each client, the standard error of the site difference (SE of $\hat{a}$wave*site) was plotted for the longitudinal analyses.

Figure 2 the standard error as a function of the number of subjects and the number of observations per subject. As subjects are added, we see the familiar shrinkage of standard errors. At first gains are greater than they are with later additions to N. In addition, differences in precision occur as a function of the number of observations per subject. When the number of subjects is small (10), a 3-wave design has a standard error of 1.4; for 7 waves, the standard error is 1.1.

Why do multi-wave designs shrink standard errors of measurement? Because they provide additional observations, as shown in Figure 3: over page.

Conclusion

Multi-wave evaluation designs add repeated measurements beyond the two measures used in pre-post designs. These multi-wave designs can be analyzed conveniently with commonly available longitudinal models (such as SAS PROC MIXED). The first few added waves

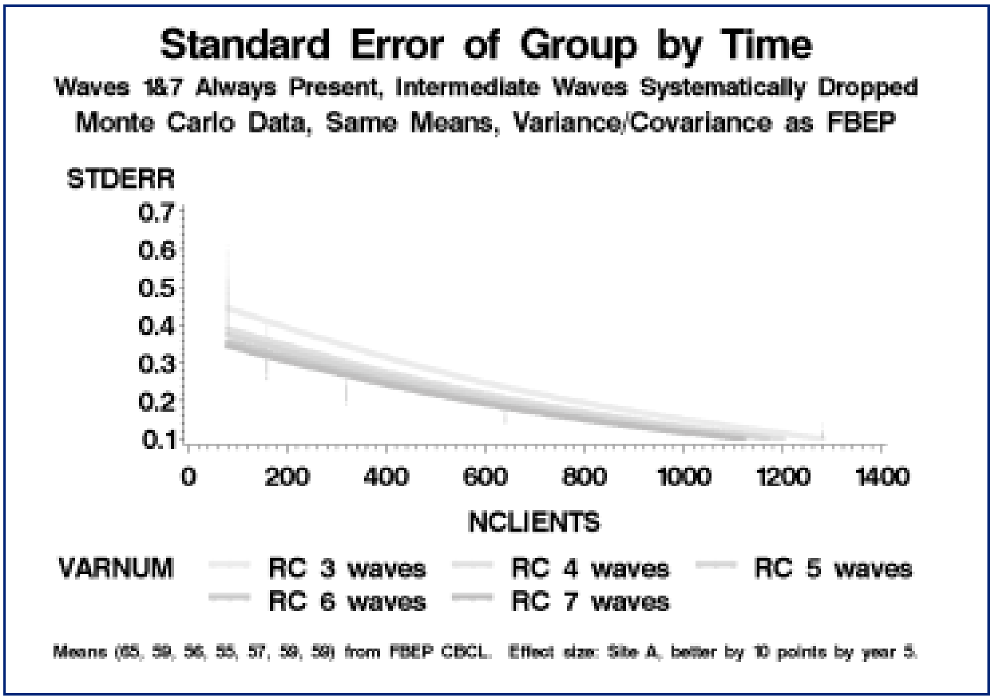


Figure 2

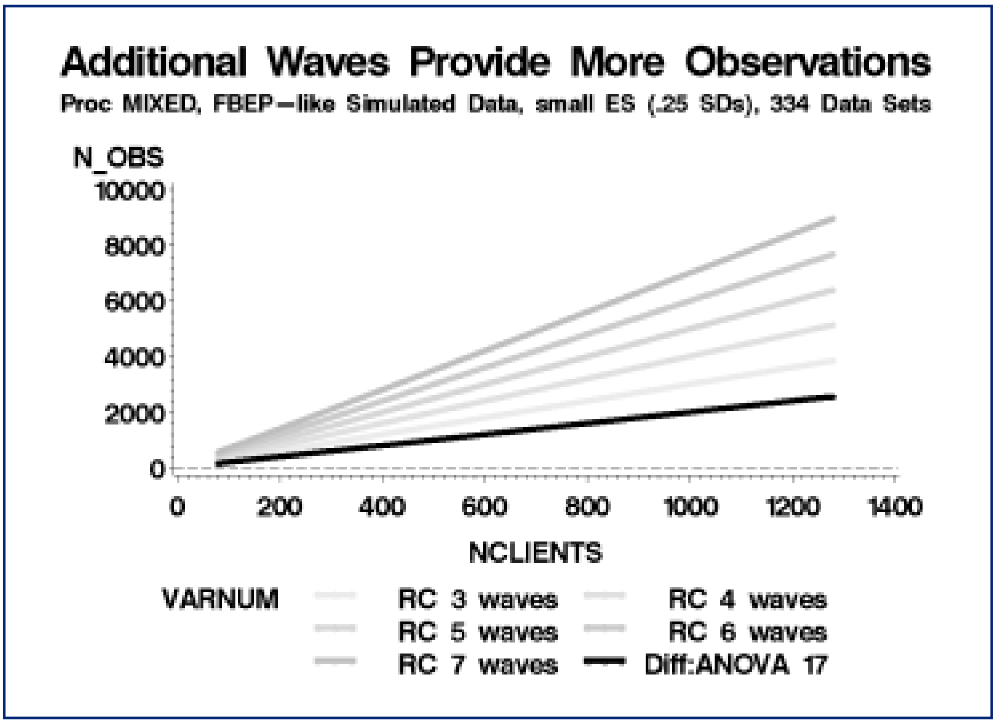


Figure 3.

beyond two increases the study's power dramatically, but by seven waves a point of diminishing returns is reached. Often it is less expensive to add a few repeated measurements to subjects already recruited than to recruit new subjects. This efficiency often makes the multi-wave design a cost-effective way to increase statistical power without simple enlargement of the study sample.

Pre-post Design is Over-parameterized.

Abstract: The pre-post design creates measurement paradoxes by using two observations to answer three questions, statistically estimating more parameters than degrees of freedom permit. Naturally contradictions and paradoxes result. Longitudinal analysis of two-wave or three-wave data suggests that three is the minimum number of waves able to answer the required questions without the problem of df = 0.

Introduction

One problem with the pre-post design is the long history of debate on two-wave change scores (e.g.,Cronbach & Furby, 1970; Edwards, Hotch, & Bickman, 1982; Rogosa, 1988). Traditionally, these arguments applied psychometric theory to pre-post data to produce infinitely debatable results (Cronbach & Furby, 1970; Gottman & Rushe, 1993; Locascia & Cordray, 1983; Lord, 1963) in an argument that was not resolved after 30 years of debate. In the present section we abandon this classical approach in favor of multi-wave longitudinal models (Gibbons et al., 1993; Rogosa & Saner, 1995) to the pre-post design to see what these more sophisticated models can show us about it.

When we ask whether mental health outcome is better in Group A or Group B, we necessarily ask three questions: a). Did the whole sample improve over time? b) Were the groups equal in severity at intake? c) Did one group improve more or improve faster than the other?

Before we can conclude “Group A improved more” we need answers to all three questions. If we knew nothing but “Group A had a better endpoint” several explanations are possible. Statistically getting answers to three questions with only two observations sounds too good to be true, like a statistical free lunch. That should serve as a warning.

Method & Results

The present study, unlike the power analysis above, uses real client scores, not simulated data. Table 1 shows the parameter estimates from random coefficients longitudinal analysis done with SAS PROC MIXED (Little, Milliken, Stroup, & Wolfinger, 1996). This is the analytic model recommended by (Gibbons et al., 1993) for longitudinal psychiatric data. Reading down the list of interpretations of the parameters shows that this procedure asks and answers the essential questions that one brings to a two-group outcome study.

Table 1 shows a longitudinal result. Interpretation of each of the term shows how well this model answers three questions:

1. Did the whole sample improve over time?
2. Were the groups equal in severity at intake?

3. Did the experimental groups improve more or improve faster?

Scores in Tables 1and 2 describe 350 children's Child Behavior Checklists (CBCLs, Achenbach, 1994). They are real scores from Stark County (Bickman, Summerfelt, Firth, & Douglas, 1997b; Bickman, Summerfelt, & Noser, 1997c), not simulated CBCLs.

The rows in Table 1 are numbered to correspond to the three questions, and the estimates tell a clear story about the outcome: At intake, the whole sample had a CBCL of 66.2, and the whole sample improved about 7 points per year, a significant improvement (p < .001). The experimental groups were 2.7 points lower at intake, but this difference was not significant (p = .11), so we consider the groups were about equal at intake. Over time, the experimental group improved 2.1 points per year more than the control group, but this difference was not significant (p = .09). Conclusion: the average client improved, and improvement was about equal in the two groups.

Table 1.—Results of random coefficients outcome analysis with 3-waves (real data)

Effect	$\hat{a}$	Std Error	DF	t	Prob.	Interpretation
0. Intercept	66.20	1.12	173	59.00	<.001	Grand mean > 0
1. Time	-7.37	0.81	173	-9.14	<.001	Average case improves
2. Group	-2.70	1.69	175	-1.59	0.11	Comp not lower at intake.
3. Group by time	-2.10	1.23	175	-1.71	0.09	Comp did not improve more.

Table 2. — Random coefficients outcome analysis with 2-Times (real data)

Effect	Estimate	Std Error	DF	t	Prob.	Interpretation
1. Intercept	67.30	1.12	173	60.06	<.001	Grand mean > 0
2. Time	-7.62	0.80	173	-9.49	<.001	Average case improves
3. Group	-2.36	1.69	0	-1.40	**==>??<==**	Comp lower at intake??
4. Group by time	-2.30	1.21	0	-1.90	**==>??<==**	Comp improves more??

The same data were re-analyzed using only the pre-test and post-test scores, to produce the results in Table 2 above. In the pre-post version, the degrees of freedom are zero for the questions of greatest interest. If we have two points and ask the three questions above, we would be asking for more information than available degrees of freedom permit. Estimates of uncertainty are impossible to make.

In the longitudinal model's results in Table 2, the question marks represent probabilities we need to know but which cannot be calculated from a pre-post design because zero degrees of freedom remain to estimate their significance. These question marks show that we can not test the significance of the results that interest us the most - were the groups equivalent at intake, and was their improvement over time different.

Part of classical psychometric theory is “Lord's paradox” (Lord, 1963), which occurs when difference scores and corrected endpoints do not agree despite the fact that both are reasonable definitions of change. In a pre-post design, in which treatment groups differ on a pretest measure, we showed above that controlling the pretest and analyzing the difference score yields different conclusions. Table 2 might be paraphrased as “Lord's warning,” if we have two points and three questions, there will be zero degrees of freedom. If we press ahead disregarding the warning, contradictory results are possible just as they are if we divided by zero.

Discussion

Experimental designs are often debated without considering the analytic models used to analyze results. An intelligent statistical model, such as random coefficients regression, provides valuable information about the pre-post quasi-experimental design. To answer the three questions we need more than two observations per subject.

Individual Post-tests: Difficult to Predict

Abstract: To evaluate individual progress, one asks whether the observed outcome at post-test was better or worse than expected for a given type of case. Using FBEP data in which we predict endpoint we found that the uncertainty surrounding predicted outcome to be so large that it would be impossible to know whether an individual client's score was better or worse than expected. If we cannot evaluate individual outcomes as expected or unexpected from pre-post data, an alternative would be concurrent monitoring, i.e. following each client with more than two repeated measurements.

Introduction

To determine whether the clients of a given service provider have good or bad outcome, we must compare a child's status on follow-up to what is expected for cases of their type. If a given provider had results that were better or worse than expected, these unexpectedly good or poor outcomes might be noted and used for quality assurance practices, or for choosing the best providers for a panel.

Good outcomes must be adjusted for "case mix," i.e. difference in caseload among providers. Without such adjustment, providers willing to work with more difficult cases might be punished, and providers clever enough to "cherry pick" easy cases would be rewarded. A medical analogy would be calling a doctor excellent when the common cold ran its course, or blaming an oncologist for having so many patients who die of cancer. This potential confounding of case mix with outcome is the reason why observed outcomes must be compared with outcomes expected for a given type of case.

When we say "expected" we might as well say "predicted. If case mix adjusted outcomes for individuals are to be useful in the field, predictions must be precise enough so that it is possible to recognize whether the outcome of a given case was better or worse than expected. The present appendix uses children's mental health data from the Fort Bragg Evaluation Project (Bickman, 1996; Bickman et al., 1995) to see how feasible it is to recognize whether a given individual's outcome is better or worse than expected.

Method

For the present study, the CBCL total score was used because it is a quantitative score from a well-designed assessment that has been widely used and researched. Using 7 measurements repeated over 5 years, the present appendix investigates how well we can predict the six endpoints from facts known at intake. These facts include 71 client characteristics, and the intake CBCL score.

Results

Table 3 shows variance accounted for by stepwise regression based on the model shown in equation 1.

Endpoint = F(intake CBCL, 26 diagnoses, 59 case characteristics) [Eq 1

In this overview, we skip a detailed examination of the predictor list and look directly at the success of prediction, shown in Table 3.

In the table we see that R2, the variance accounted for, starts at 48% at six months, declining to 28% when we predict outcomes five years later. These R2 are large (Cohen, 1988; Cohen, 1992) and typical for the field (Achenbach, Howell, & McConaughy, 1998).

Table 3—Prediction Intervals for N = 635 Treated Children

Years (1)	R-squared (2)	Prediction Interval (3)
0.5	0.48	34.1
1.0	0.39	37.1
1.5	0.31	41.0
3.0	0.35	37.9
4.0	0.31	40.1
5.0	0.28	36.4

Notes:

1. Prediction intervals computed for cases with missing CBCL endpoint based on cases present.
2. The predicted Y is the endpoint, not the more difficult problem of predicting change (slope, residual, or difference.) Using CBCL as both Y and an X (as Achenbach did) makes R2 larger.

Statistical regression procedures[1] produce two kinds of confidence intervals:

Confidence interval for a whole sample. A 95% confidence interval tells us the range within which we are 95% sure the “true” regression line lies. This kind of confidence interval tells us the accuracy of parameters for a whole sample but tells us nothing about our ability to predict the future outcomes of a given case.

Prediction interval for an individual client. Another kind of confidence interval is called a “prediction interval,” which tells us the uncertainty when we estimate the future value of

---

[1] *For example SAS PROC REG (ordinary least squares regression).*

estimate for a particular individual. For example, if John Jones expected future CBCL score is 60 and the 95% prediction interval has a range of 10, we might say that his expected score is 60 ± 5.

Prediction intervals of 95% confidence are shown in Table 3 column 3. They are very large. To see how large, let's consider prediction at six months, as shown in Table 4.

Table 4—Prediction of CBCL at six months from facts known at intake

Variable	Label	N	Mean	Std. Dev.	Min.	Max.
LO_PRED	Lower Bound of 95% C.I. (Individual Pred)	635	41.35	8.12	14.65	62.00
UP_PRED	Upper Bound of 95% C.I. (Individual Pred)	635	75.48	8.14	49.03	96.31
INDIV	Individual prediction interval	635	34.14	0.09	34.03	34.53

Notes:
CBCL: CBCL Total Problem T-score
Facts known at intake: CBCL at intake + 26 diagnoses + 59 child and family characteristics

As shown in Table 4, the average expected score on the CBCL at six months ranges from 41.35 to 75.48 (with 95% confidence), a range of 34.14 points (about 3.4 standard deviations). This large range suggests our prediction is statistically significant, tells us quite a lot about groups of clients, but that it is not clinically useful for predicting the endpoint of a given individual client. It is like predicting "Johnny's grade will be between A and F inclusive."

It is not useful in the individual case because of the huge zone of uncertainty of over 3 standard deviations. Thus our prediction for a large group may be fairly good (R2 = 48%), but nowhere near good enough for the accurate prediction of one individual's fate.

The six charts in Figure 4 show these intervals for the six follow up intervals. The main point of these six charts is that we cannot predict with any certainty what score is expected for a given individual patient. Unless unbelievably high predictive correlations (r > 0.90) were available. Predicting the mean of a group (e.g. N = 100) is much easier than predicting the score of a single individual.

Discussion

The large interval of uncertainty for individual prediction does not support the use of predicted individual endpoints in quality control for individual cases. If we cannot predict an individual's future with any certainty, how can outcome best be monitored? At the Vanderbilt CMHP we are now investigating two more feasible methods both of which offer more immediate payoffs than prediction-based evaluation schemes.

Concurrent monitoring. If we cannot predict the future with precision, we can evaluate outcomes better with a more detailed view of each child's progress to date. With frequently repeated outcome measurement (e.g. weekly), two new ways to evaluate outcome appear that would not be possible with pre-post data.

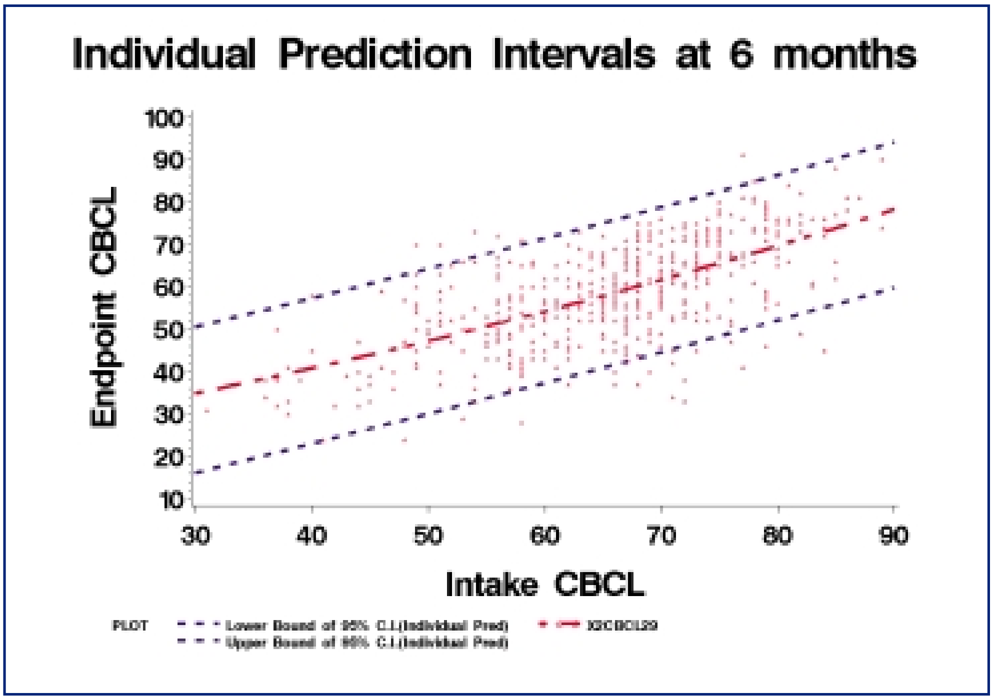
Individual growth curves: With more than four repeated observations per client, it becomes meaningful to speak of each client's growth curve parameters, e.g. using a random coefficients time-regression model or Bryk & Raudenbush's (1992) hierarchical linear models. These two level models reconcile growth curves of individual clients with the average growth of the group. Using these advanced statistical models, we have pilot tests of a way to "flag" cases when the child's status falls outside that child's 95% interval of confidence. Since this model uses all data points for a given child, higher R2 can be achieved by more frequent measurement, making the individual prediction intervals shrink.

Conclusion. At the present time we believe that outcome evaluation methods based on prediction of a given client's future status have face validity, but for technical reasons are not feasible at the present time. Currently available predictive R2 in the 40% to 60% range are not sufficient for accurate individual prediction.

At the present time concurrent monitoring (Howard, Moras, Brill, Martinovich, et al., 1996) meaning a more detailed view of each client's individual progress through multiple repeated measurements at short time intervals (e.g. weekly), analyzed with currently available longitudinal statistical models offer a useful and feasible alternative to prediction-based schemes. If we can't predict the future very well, we can examine the past and present in greater detail, and study time as the important variable that it is, rather than using pre-post designs and statistics reduce time to a constant.

Even if we cannot predict with certainty a client's future progress, we can improve the evaluation of outcomes by a more detailed analysis of the client's progress to date. An example of the statistical details for doing this appear in the last unit of this appendix.

Figure 4—Prediction of individual scores: wide intervals of 95% confidence



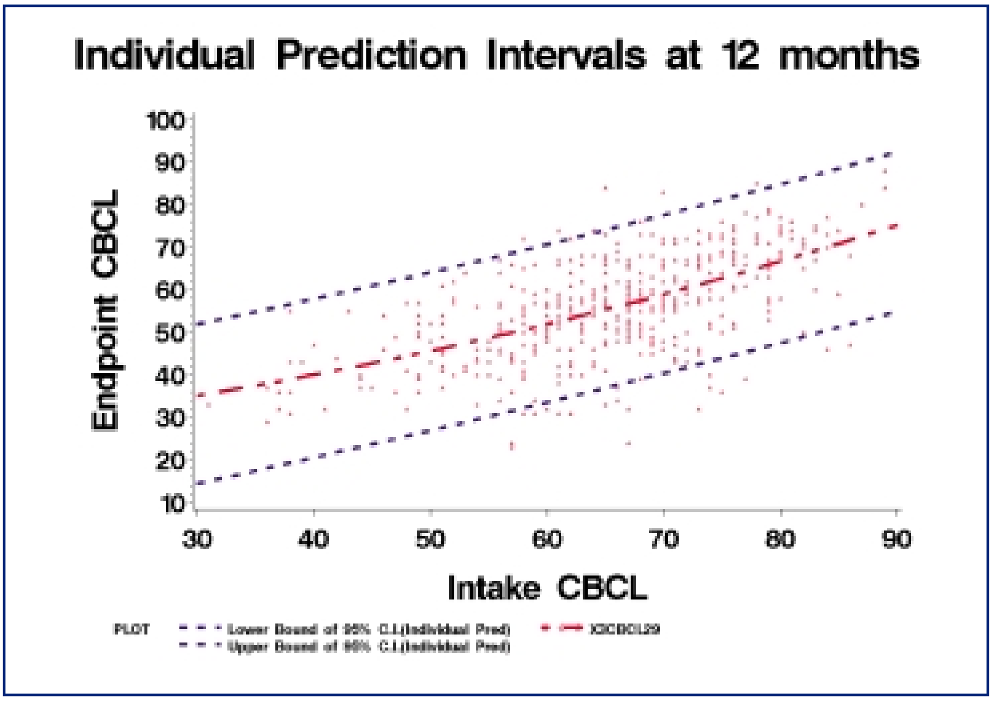
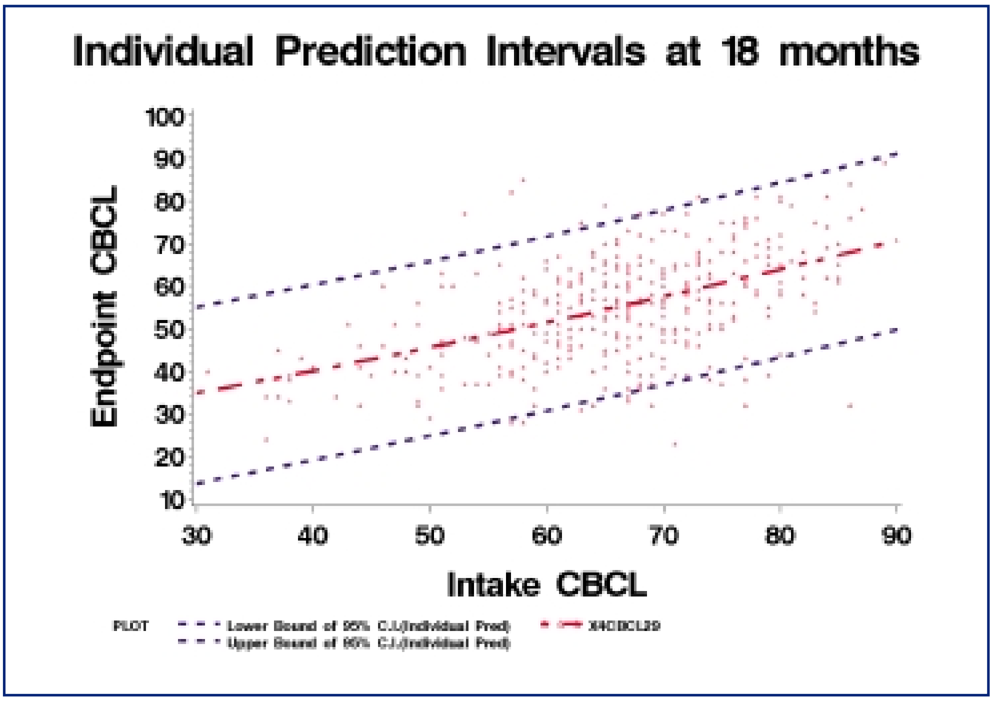


Figure 4—Prediction of individual scores: wide intervals of 95% confidence cont'd



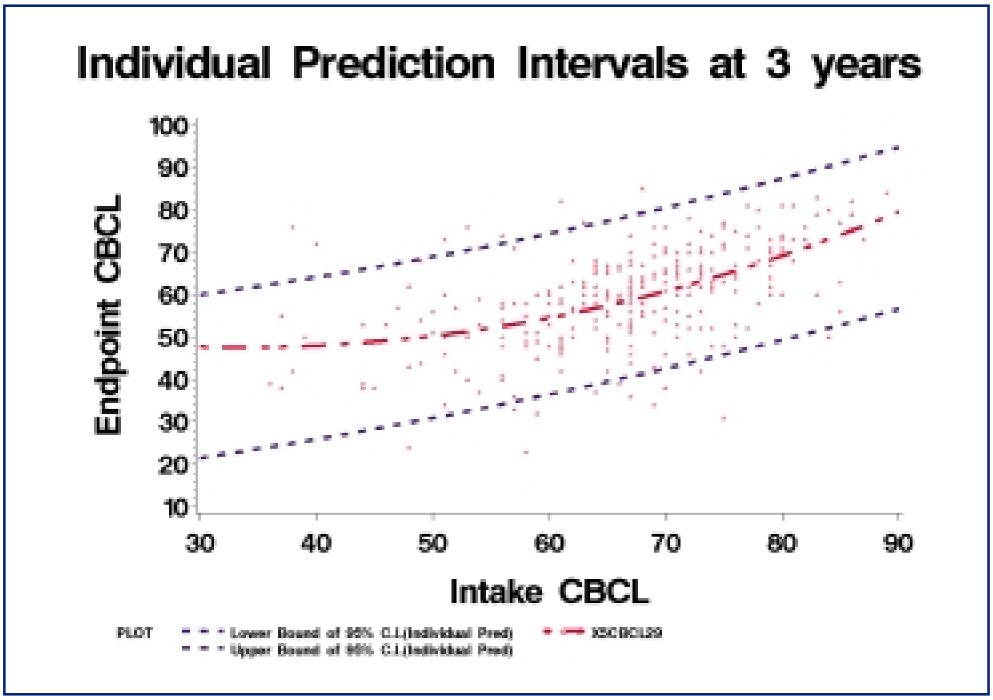
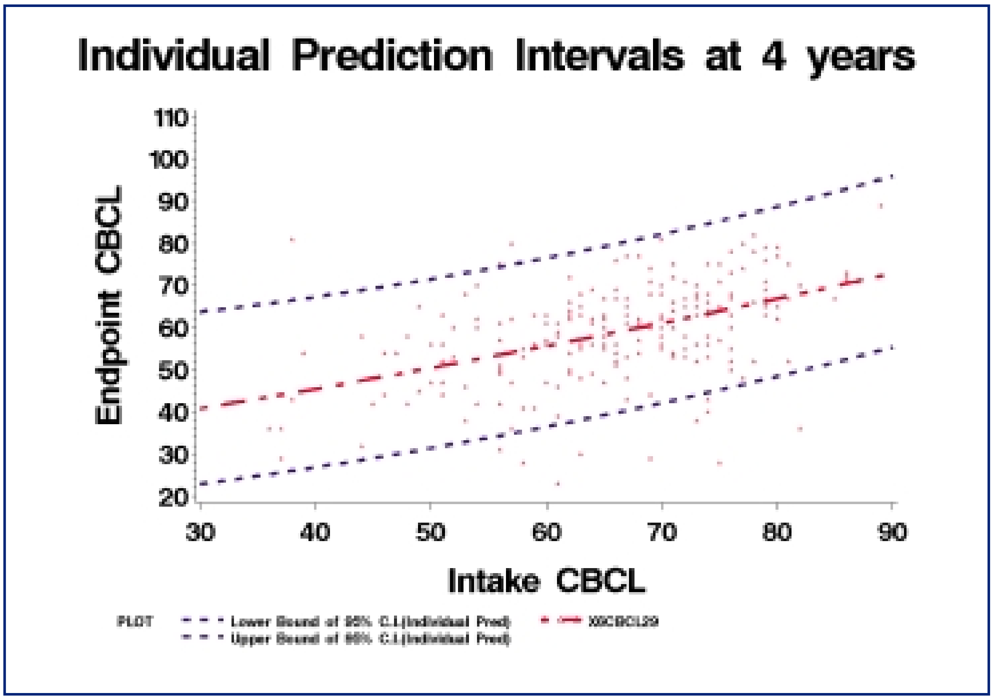
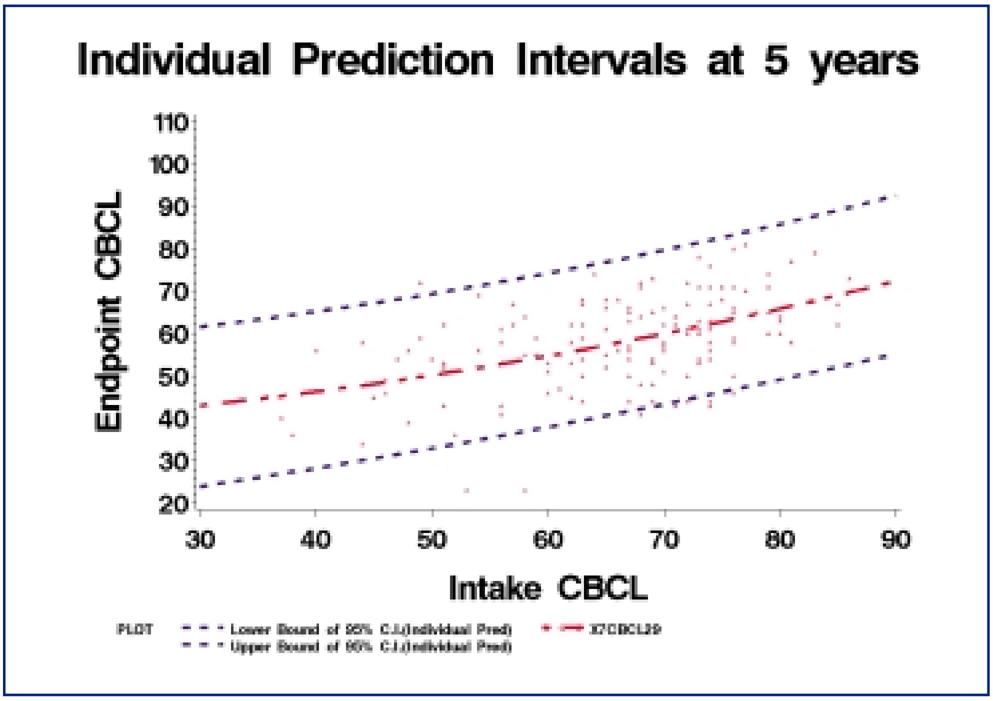


Figure 4—Prediction of individual scores: wide intervals of 95% confidence cont’d





Individual Progress Charts with Warning Flags[2]

Abstract: An analytic model for continuous outcome monitoring is proposed. It is based on a longitudinal model recommended for psychiatric data and can be implemented with commonly available software (e.g., SAS). The model graphically tracks a client's individual progress based on the child's own intercept (severity at admission) and rate of improvement (slope in points per month). The model presents the client's actual scores as well as a linear model for all the client's scores to date. When a new score appears that is outside the range of 95% confidence for that client, the progress chart is marked with a warning flag. The procedure produces progress charts (Cartesian timelines) to inform clients and clinicians of the progress of a case over time.

Introduction

The present section introduces a computer-generated client chart that shows client outcome on a key variable, such as CBCL Total. The proposed chart offers several advantages over measures of client change now in common use. It is longitudinal, not limited to naive pre-post change scores, and the method can accommodate any number of repeated measures per child without programming intervention. Missing waves of data require no imputation, nor are incomplete cases dropped; the longitudinal statistical model uses whatever data are available. There is no requirement that cases be measured at set intervals, e.g. weekly; times can vary within and across individuals. The output is graphic, a client literal chart simple enough for parents, clinicians, and administrators to understand. In addition to simply charting each child's individual outcome curve, the chart warns when a child's progress is significantly worse than expected (based on that child's progress and on the general pattern of children in the sample). The proposed method uses a hierarchical statistical model that combines two levels of data (group or agency level and child level) for the most precise estimates. The method is exportable through SAS, and can be done with a few hundred lines of programming code.

A pre-post endpoint model doesn't fit clinical practice, where clients are seen repeatedly with many sessions across varying lengths of time, so that a longitudinal description of outcome is needed.

The need for a real patient chart with warning flags

When program evaluators gather their scales and scores, clients, clinicians, administrators, and policymakers often ignore the results, many times because client outcomes do not appear in clear and simple form they can quickly assimilate. In the present example, a key outcome (CBCL Total) for each child appears as a chart or timeline simple enough to be understood by clients with no background in statistics or research.

---

[2] *An earlier draft of this idea appeared in an unpublished report (Bickman, Lambert, Northrup, Salzer, & Summerfelt, 1997a).*

The proposed charts, two examples of which appear below, have three features:

- A timeline of the child's CBCL scores
- A curvilinear longitudinal model of the child's timeline of CBCL scores
- A warning flag if the observed CBCL is outside a 95% interval of confidence around the model (i.e. unexpected deviations are flagged).

Method

The present demonstration used all children in the FBEP with five waves of nonmissing CBCL scores to demonstrate the concept. Analysis was done with SAS PROC MIXED using a longitudinal random coefficients model. In this model each subject has their own individual parameters; group level parameters are calculated as well. To make the estimates as precise as possible at both levels, SAS PROC MIXED refines solutions iteratively until the overall solution has the best overall fit at both the individual and group levels. The code for this analysis[3] follows:

Equation 1. A Longitudinal Model of Outcome

```
proc mixed ;
        class site Case_ID;
        model cbcl = site months
                site*months months*months
                site*months*months/chisq solution predicted;
                random intercept months /
                type = unstructured subject = Case_ID;
                make ‘predicted’out = p noprint;
```

This analysis creates a model with linear and quadratic time, the latter being necessary for a curve rather than straight line. Individual timelines were made by passing results from the outcome analysis to SAS PROC GPLOT. This approach is convenient with data sets of any size, even tens of thousands of clients.

Results and Discussion

Individual Progress Charts with Warning Flags: Two examples

Of the hundreds of charts that were generated, two examples were selected for the present report. The first example appears in Figure 5.

---

[3] *The variable "site" refers to the FBEP Demonstration vs. Comparison sites. Any grouping variable could be used, such as clinic A vs. clinic B, C, D . . ., or clinician A vs. clinician B.*

Figure 5A—Individual Report without Warning Flag

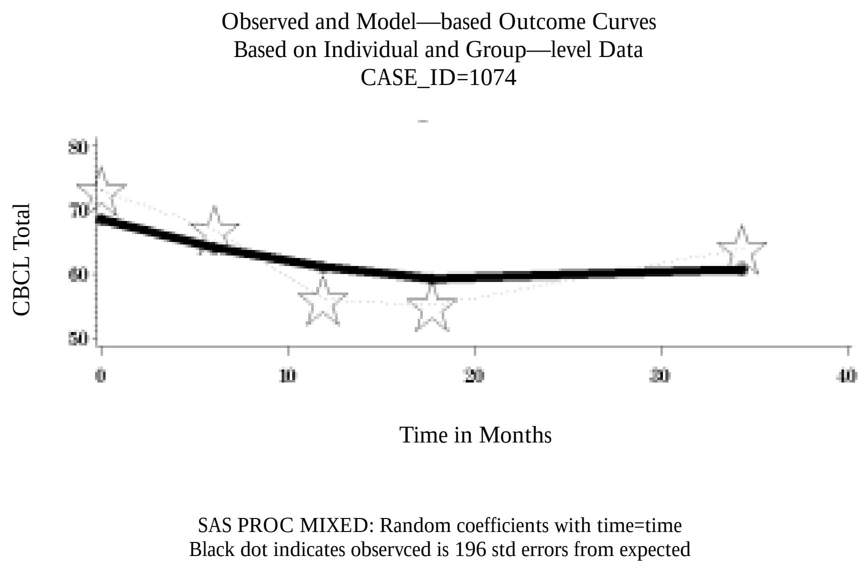


Figure 5B. Individual Report without Warning Flag

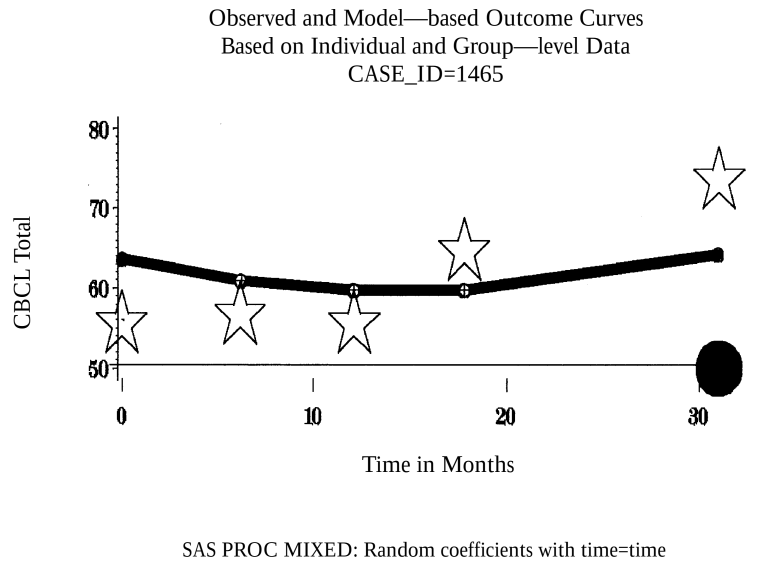


Figure 5—Individual Report without Warning Flag

In the first example, a child’s actual scores appear as large stars around child #1074’s individual response model (the heavy line). This case shows a typical pattern in the FBEP, namely initial improvement followed by diminishing returns over time. In the first example, there were no observations with significant (p < 5%) departure from the individual child's model; thus we may say that nothing unexpected occurred. For child #1074 the heavy black line is the best summary of the child's outcome over five repeated measurements.

The second example, shown below in Figure 5b, illustrates the automatic warning flag. The black dot on the X-axis is the warning flag at Wave 5. The warning appears because the Wave 5 CBCL value for child #1465 was more than 1.96 standard errors from the value expected in the child's individual time line model.

Parents and clinicians of case #1465 should already know about the unexpected deterioration at Wave 5; if they do not, it is something they should investigate. Administrators and quality assurance staff might ask did the treatment team document this problem in the chart? Is there a plan that attempts to cope with the child's unexpected deterioration at Wave 5?

While the two figures were hand selected for this report, it is convenient to conduct SAS in batch mode to print hundreds or thousands of charts automatically using a common desktop computer. Charts could be printed for all clients, or selected ones, e.g. those with warning flags. Program evaluators could use the group-level parameters of the longitudinal analysis to evaluate outcomes of different clinicians, clinics, or years. Clinicians and parents could use such charts to monitor a child's overall progress at a glance

References

Achenbach, T. M. (1994). Child Behavior Checklist and related instruments. In M. Mark Edward (Ed.), *The use of psychological testing for treatment planning and outcome assessment*. (pp. 517-549): Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US.

Achenbach, T. M., Howell, C. T., & McConaughy, S. H. (1998). Six year predictors of problems in a national sample: IV. Young Adult Signs of Disturbance. *Journal Americaln Academy of Child and Adolescent Psychiatry*, 37(7), 718-727.

Bickman, L. (1996). A continuum of care. More is not always better. *American Psychologist*, 51(7), 689-701.

Bickman, L., Guthrie, P. R., Foster, E. M., Lambert, E. W., Summerfelt, W. T., Breda, C. S., & Heflinger, C. A. (1995). *Evaluating managed mental health services: The Fort Bragg experiment*. New York, NY, US: Plenum Press.

Bickman, L., Lambert, W., Northrup, D., Salzer, M., & Summerfelt, T. (1997a). *A Draft Instrument with Empirical Foundations* . Nashville: Vanderbilt University.

Bickman, L., Summerfelt, W. M. T., Firth, J. M., & Douglas, S. M. (1997b). The Stark County Evaluation Project: Baseline results of a randomized experiment, *Nixon, Carol T. (Ed); Northrup, Denine A. (Ed); et al. (1997). Evaluating mental health services: How do programs for children 'work' in the real world? Children's mental health services, Vol. 3. (pp. 231-258). Thousand Oaks, CA, USA: Sage Publications, Inc. 305 pp.SEE BOOK* .

Bickman, L., Summerfelt, W. T., & Noser, K. (1997c). Comparative outcomes of emotionally disturbed children and adolescents in a system of services and usual care. *Psychiatric Services*, 48(12), 1543-1548.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA, US: Sage Publications, Inc.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and Quasi-experimental Designs for Research*. Boston: Houghton Mifflin Company.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. (2nd ed.). Hillsdale, N.J.: L. Erlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.

Cronbach, L. J., & Furby, L. (1970). How we should measure "change": Or should we? *Psychological Bulletin*, 74(1), 68-80.

Diggle, P., Liang, K.-Y., & Zeger, S. L. (1994). *Analysis of longitudinal data*. Oxford and New York: Clarendon Press ; Oxford University Press.

Edwards, J. D., Hotch, D. F., & Bickman, L. (1982). Measuring children's health-related attitudes and knowledge. *Applied Social Psychology Annual*, 3, 249-277.

Gibbons, R. D., Hedeker, D., Elkin, I., Waternaux, C., et al. (1993). Some conceptual and statistical issues in analysis of longitudinal psychiatric data: Application to the NIMH Treatment of Depression Collaborative Research Program dataset. *Archives of General Psychiatry*, 50(9), 739-750.

Gottman, J. M., & Rushe, R. H. (1993). The analysis of change: Issues, fallacies, and new ideas. *Journal of Consulting and Clinical Psychology*, 61(6), 907-910.

Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z., et al. (1996). Evaluation of psychotherapy: Efficacy, effectiveness, and patient progress. *American Psychologist, 51*(10), 1059-1064.

Khattree, R., & Naik, D. N. (1995). *Applied multivariate statistics with SAS software*. Carey, NC: SAS Institute.

Little, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). *SAS System for Mixed Models*. Cary, N.C.: SAS Institute.

Locascia, J. J., & Cordray, D. S. (1983). A Re-Analysis of "Lord's Paradox.". .

Lord, F. M. (1963). Elementary models for measuring change. In C. W. Harris (Ed.), *Problem in Measuring Change* (pp. 259). Milwaukee: University of Wisconsin Press.

Rogosa, D. (1988). Myths about longitudinal research. In R. T. C. W. M. S. C. R. K. Warner Schaie (Ed.), *Methodological issues in aging research*. (pp. 171-209): Springer Publishing Co, Inc, New York, NY, US.

Rogosa, D., & Saner, H. (1995). Longitudinal data analysis examples with random coefficient models. *Journal of Educational and Behavioral Statistics, 20*(2), 149-170.

APPENDIX 7

MEETING THE CHALLENGES IN THE DELIVERY OF CHILD AND ADOLESCENT MENTAL HEALTH SERVICES IN THE NEXT MILLENNIUM: THE CONTINUOUS QUALITY IMPROVEMENT APPROACH[1]

Leonard Bickman and Kelly Noser
Vanderbilt University

This is a time of major worldwide upheaval and change in the mental health field. In the United States the introduction of managed care, where private for-profit companies regulate care, is a major source of change. Economic rationalism has compelled many governments to examine all services, but especially human services, to see if they are providing value for money. Many of these forces have focused on concerns about the quality of mental health services (Bickman & Salzer, 1997; Nash, 1995; Zieman, 1995). Some predict that quality will surpass cost control as the primary task of health care management (McGlynn, Halfon, & Leibowitz 1995).

These trends, both positive and negative, prompt change in the way our service delivery organizations operate in the next century. This focus on quality and cost effectiveness should have great appeal to consumers since it moves them not only to demand better access and more services but also, to insist that services be effective. Policy makers and managers should find this emphasis attractive since it is designed to provide more cost-effective services. Clinicians should welcome this change since it provides them with tools to sharpen their clinical skills. However, we need an approach and direction to deal with the challenges of providing high quality, cost effective services. The approach that we are advocating is what has commonly been known as continuous quality improvement or CQI. Briefly, CQI requires that we understand the linkages between process of care and outcomes, that we systematically collect data on these linkages (Deming, 1982; 1986), and that we create an atmosphere of change at all levels of the organization that acts to implement changes in daily protocols based on the observed linkages (Dickens, 1994)

We are first going to review the forces that are propelling the field towards change so that we can have a better understanding of the context of this change for the 21st century. Next, we are going to describe CQI, which we believe can help us through this turbulent period and improve services. But, we also will point out the major obstacles we face in applying CQI to

---

[1] *Keynote address delivered to the Third National Conference on Child and Adolescent Mental Health, July 1998, Sydney, Australia. Professor Bickman can be contacted at Bickman@home.com. More information can be obtained at http://www.vanderbilt.edu/VIPPS/CMHP/cmhphome.html*

mental health so we do not adapt this approach naively. Finally, we will suggest some ways we can overcome some of these obstacles. But before we go into any more details on CQI we want to describe why such an approach is needed

Forces Moving Us Towards a CQI Approach

The CQI approach is motivated by four themes. First, research we have done has shown that the dominant reform movement in child and adolescent mental health services, systems or continua of care, has not had the hoped for impact on outcomes (Bickman, Noser, & Summerfelt, 1994; Bickman, 1996). Second, the recent awareness that treatment research that has been primarily laboratory based appears not to generalize to community settings where care actually takes place (Weisz, J. R., Weiss, B., Han. S.S., et al. ,1995). Third, the new emphasis on accountability and outcome measurement especially if payment is based on consumer improvement (Markson & Nash, 1995; Panzer, 1995). Finally, many professional organizations and accrediting institutions are developing practice standards (and if followed by practitioners should result in better client outcomes (Donabedian, 1986). We believe these complimentary forces support the introduction of CQI.

The Failure of Systems of Care to Demonstrate that Client Outcomes are Improved.

During the past several years, children's mental health field has been heavily influenced by the belief that an improved system of care that included better coordination, more family involvement, a wide range of services, and case management, would lead to improved child outcomes. (Behar, L, 1988; Stroul & Friedman, 1986; Saxe, Cross , Silverman, et al. 1987; Friedman, Duchnowski, 1990)

For example the U.S. Center for Mental Health Services has a 70 million-dollar a year program designed to enhance local systems of care. We have conducted two extensive longitudinal evaluations to examine the effectiveness of carefully designed systems of care. These studies are the Ft. Bragg and Stark County studies (Bickman, Noser, & Summerfelt, in press; Bickman, 1996). Our results from these studies show that while these two systems of care improved access to services they had no additional effect on child level clinical outcomes. Furthermore, they greatly increased costs (Bickman, Guthrie, Foster, et al. 1995; Foster & Bickman, 1998). Children and youth did improve but it did not matter whether or not they received the enriched system of care services. These results held for five years after children started treatment. Our results are not isolated. Reviews of innovative system level interventions such as case management have also failed to provide good evidence of effectiveness (Burns, Farmer, Angold et al. 1996; Rivera & Kutash, 1996.). Given the findings that improved system-level changes and many innovative system level innovations are not sufficient to improve child outcomes, policy makers, consumers, researchers and practitioners need to focus on other areas that have the potential to deliver better mental health services to children and their families.

The Lack of Evidence of the Effectiveness of typical treatment.

Until very recently, it was thought that there was powerful scientific evidence of the effectiveness of treatments such as psychotherapy. In the adult mental health field, thousands of studies demonstrate that psychotherapy is effective in treating many disorders (Smith, 1977). While the evidence in the child field is not as extensive, it still is quite convincing (Weisz, et al., 1995). However, there was little awareness of the fact that almost all of these studies were conducted in very controlled university based clinics and not in real-life situations involving the kind of children usually referred for treatment. Moreover these studies are not conducted with typical clinicians with varying caseloads, expertise, supervision, and training. Efficacy research addresses the question of what are the effects of treatment under optimal conditions. Efficacy research may be helpful in the developing and evaluating the early stages of clinical treatment protocols. Once such protocols have proved successful, they can be (but are usually not) used in training and continuing education of new and seasoned practitioners. These treatment protocols are seldom evaluated in everyday clinical settings (Hoagwood, Hibbs, Brent & Jensen, 1995).

In efficacy research, the study sample is carefully screened so that each participant meets the one diagnostic criterion for the specific disorder to be treated by the protocol. Co-morbid consumers, that is clients with more than one diagnosis, are usually not included in these studies. In contrast, in effectiveness studies, the sample generally represents wide variety children referred to mental health services. While children in efficacy research studies tend to have a single diagnosis, those in effectiveness studies are more likely to have multiple diagnoses. Service providers participating in efficacy research also differ greatly from those in effectiveness studies. Those in efficacy research are generally in academic settings, for example graduate students in clinical psychology, while those in effectiveness research are in more practice-oriented environments, for example, private practice and community mental health clinics. Moreover, the therapists in efficacy trials are rigorously trained in the specific treatment that is being tested. The therapists are carefully supervised, often videotaped, to ensure that the treatment protocol is followed. This is not the situation in community settings.

There are not many studies of child and adolescent treatment in the real world. However meta-analytic reviews of those studies have concluded that there was no evidence of effective treatment for children and adolescents in real-world settings (Weisz, et al., 1995). In addition, we recently completed an outpatient dose-response study that found no evidence that outcomes improved with more treatment (Salzer, Bickman & Lambert, in press). These results suggest that the lack of a dose-effect response reflect the ineffectiveness of usual services. That is, since there is no difference in outcomes after receiving one or two sessions or fifty sessions of outpatient treatment, it may be that treatment is not effective. Please note that we are not saying that there is proof that services are ineffective. We are saying that there insufficient scientific evidence to support the contention that services are effective. It is not possible to prove that services everywhere are ineffective. We believe that there are effective services and clinicians, but we have not yet identified them through research. However, the lack of evidence must be taken seriously given the other forces that we will describe next.

The Accountability and Outcome Measurement Movement

Some describe the emphasis on outcomes as outcome mania. Others see the new importance of outcomes as a positive force. Whatever the perspective it is clear that most believe there is added significance of outcome measurement (see the recent special issues *of Behavioral Health Management*, Vol. 15, #3 and *New Directions for Mental Health Services*, Fall 1996, #71). This emphasis is recognized in Australia in the National Standards for Mental Health Services (1997). In the first page of the overview to the Standards it states “ The standards are outcome oriented with an emphasis on the end result for consumers and carers (p. 1).” In the United States several organizations are either developing standards for outcome measurement or specifying instruments to be used in collecting information from consumers and clinicians. Several states in the U.S. are writing outcome measurement requirements into their contracts with managed care companies One leading consumer organization, the National Alliance for the Mentally Ill, has convened an outcomes roundtable to provide direct consumer input). The senior author is currently working on a project funded by Commonwealth of Australia to recommend a measurement system for both outcome and process measurement for children and youth.

Practice Guidelines.

The fourth influence on the movement towards quality is the development of practice guidelines. The American Academy of Child and Adolescent Psychiatry’s practice parameters represent the first effort devoted to the evaluation and treatment of children and adolescents (disseminated through the Journal of the American Academy of Child and Adolescent Psychiatry, supplement to Vol. 36, 1997). These practice parameters represent the current state of information on evaluation and treatment for several distinct mental health problems. Practice parameters have been developed on assessment, evaluation of child custody, and the treatment of specific disorders such as obsessive-compulsive disorder.

Dunne (1997) reports that practice parameters were originally developed to combine current empirical knowledge with clinician judgement in order to promote "best practices". The expectation is that there will be continuous development and dissemination of practice parameters, with updates that include the latest empirical data on relevant assessment and treatment issues. However, concerns have been raised about the empirical basis for the development of mental health treatment guidelines, emphasizing the need for more "real world", or effectiveness treatment outcome research (Nathan, 1998). A question that has not been raised relates to the actual use of established guidelines or parameters by practicing clinicians. We will be beginning a project with the cooperation of the American Academy of Child Psychiatrists that studies how guidelines are used by psychiatrists and eventually if they affect child outcomes. It is necessary that the field identifies and describes factors that influence clinician's utilization of the practice parameters as a component of the treatment process itself.

We see there are major international trends in mental health that are motivating concern about the quality of services and their impact on consumers. Next we will describe more about continuous quality improvement and how it can help us respond to these forces.

What is CQI?

The principles of continuous quality improvement or CQI, as applied in manufacturing and service industries, have been looked to for guidance in improving quality in the health care industry (Dickens, 1994). CQI involves the use of assessment, feedback, and application of information to improve services. A CQI strategy attempts to be proactive, rather than reactive; it is about solving problems and maintaining quality, thus relying on a continuous evaluation of processes and outcomes. CQI involves a dynamic interplay of assessment, feedback, and application of information.

The idea of providing clinicians with information, or feedback, to improve individual performance and quality of services is an adaptation of Deming's (1982, 1986) management principles. Deming, is the well-known management consultant whose teachings helped revolutionize manufacturing in Japan. Deming linked operations and outcomes data into a feedback loop that triggers a continuous improvement process for manufactured goods. The feedback strategy requires a flexible, no-threat system whereby personnel are encouraged to improve their own performance (Panzer, 1995). We want to emphasize the importance of the no threat principle by what we term the Vasa effect. The Vasa is a Swedish warship that sank on its maiden voyage, while still in the harbor in 1628 (Borgesnstam & Sandstrôm, 1995). A “test” or evaluation of its stability was conducted before it set sail that showed it was unstable. One of the possible reasons for ignoring the results of the evaluation was fear of displeasing the king who did not want any delay. More recent examples of the failure to effectively communicate the poor quality of a product or process in the face of political and financial pressures can be found in the shuttle rocket Challenger’s disaster caused by the failure of the “o” rings (Gleick, 1992).

Most problems of service quality are believed caused by external factors such as poor information, lack of knowledge, poor job design, or the complexity of the process (Berwick, 1993) and not poor quality personnel. For children’s mental health to provide such a feedback system, we must know what type of information about the treatment process is needed. This dictates that the field develops a solid knowledge about what aspects of the process lead to desired outcomes, in order to monitor them on an ongoing basis.

Another important aspect of CQI is that it enables employees at all levels to take responsibility for improving the environment in which they work and the service they provide (Glazer & Gaitner, 1995; Markson & Nash, 1995). The culture that is nurtured/encouraged in CQI structures has been referred to as a “quality circle” (Oakland, 1989) or “quality action groups”. Quality circles are comprised of many stakeholders — ranging from clients, families, and their service providers to professional specialists, and even local community patrons. Those involved in the quality circles meet regularly and voluntarily, with a mediator or supervisor to work through work-related problems and potential solutions (Dickens, 1994). Quality circles can be instrumental in developing best-practice guidelines for administrative and clinical planning. Many, however, criticize the formation and working of quality circles because they believe they are more likely to create over-enthusiastic employee motivation than anything tangible.

While many (Dickens, 1994) support the application of CQI in health care settings, there have been very few demonstrations of its implementation, much less its success, in other human service fields. Dickens (1994) notes its popularity in the following human service fields: education (Ivancevich and Ivancevich, 1992; Edwards, 1991), government (Swiss, 1992), and health care (Fried, 1992).

We would like to give you a concrete example of a very simple application of CQI in a health care setting that was recently described to one of the authors (LB). A hospital was having difficulty in getting physicians to examine the feet of patients in their physical exam. They tried the usual procedures of education and exhortation- in service lectures, items in the hospital newsletter, and individual counseling. All failed. Then they looked at the examination process in more detail and felt they could test a simple way to improve the quality of the process. In half their examination rooms they posted a large sign that read, “Patients please remove your shoes and socks”. It was reported that the physicians did examine patients’ feet - a simple and inexpensive CQI intervention.

In mental health we do not know of published examples of CQI but we would like to describe a situation the first author learned about in a workshop he gave last year in Australia. The service was a drop in center for adults who were mentally ill. The director of the center said his biggest problem was that people would come in for a few days and not return. A key part of CQI is having in place a data system that provides important information and thus allows the manger to experiment at almost no cost. In this case the manager said that the center always collected attendance data. What the manager lacked was an intervention that would increase attendance. It was suggested that he might want to try to contract with new attendees and specify a commitment to a specified number of days. He could try that with the next 20 new consumers and compare their attendance with the earlier attendance data to determine if the intervention was effective. If a data system is already in place, one can learn how to improve services at almost no cost. But it takes curiosity and discipline.

Let us provide one more example of CQI that we plan as research. We have proposed that a key clinical process is the therapeutic alliance between the clinician and the consumer. This is the emotional bond and agreement on the purpose of treatment. This alliance is considered important in most treatment modalities. We suggested that the strength of the alliance be assessed on a regular basis and be fed back to the clinician with suggestions for improvement if the alliance was not sufficiently strong. In contrast to the two previous examples, this would not be a special study but an ongoing process that could be built into treatment. Of course, before that could occur we would need good measures of alliance, interventions that could affect alliance, strong evidence that improving alliance improves consumer outcomes and acceptance by clinicians, managers, and consumers.

The Need for Process and Outcome Data

We have reviewed four factors that are motivating the mental health community to be more concerned about measurement. We have suggested that CQI can be helpful. But what do we need to measure? Is knowledge of outcomes sufficient to improve services? We want to stress the importance of being able to measure both outcomes and important clinical processes. However, recent articles and studies supporting the importance of outcome monitoring have not considered the possibility that true quality improvement depends not only on outcomes measurement but also on refining treatment based on data that relates treatment processes to outcomes.

Past research on quality of care has emphasized the need to consider process and outcome in measuring quality of services (Donabedian, 1980). We must understand the complex treatment processes of mental health care before we can improve them (Berwick, 1989). We need to link processes with outcomes. Our efforts are inadequate if they fail to establish links between key aspects of care and the outcomes produced (Bickman & Peterson, 1990). Outcome information alone will not improve services unless accompanied by reliable measures of the processes that produced them. These processes have to be validated as quality indicators by establishing causal links with outcomes.

A key component of CQI is the timely feedback of accurate information. Do we need to have standardized and psychometrically sound instruments to accomplish this? We know that clinicians do receive informal feedback. However, research has shown it is difficult for clinicians to decipher and interpret information about clients using the relatively unstructured and unsystematic observations that occurs during treatment (Smith, 1988). In fact, most problems with quality may be caused not by incompetent clinicians, but by external factors such as poor information, lack of knowledge, and the inherent complexity of the therapeutic process (Berwick, 1993; Laffell & Blumenthal, 1989).

The feedback system could influence clinical outcomes by providing information that enables clinicians to make more informed decisions, enhance the treatment process, and thus improve outcomes. Without a systematic and valid feedback system, clinicians may not have adequate information about their clients, and thus may be unable to determine whether their decisions are correct. Structured feedback should allow clinicians to evaluate their decisions against valid data about their clients; they can then alter their decisions if they receive information that appears to indicate that change is necessary.

Although the potential for outcomes monitoring and feedback to improve the quality of mental health services is recognized, there is little empirical evidence that feedback improves quality of care. However, there is some evidence that these systems can be implemented in community settings. In Australia, Michael Sawyer, Sarria, & Baghurst (1991) developed a computerized monitoring system to use for assessment purposes. In the U.S., Ken Howard and his colleagues have used a concurrent monitoring system with adults in which client-specific process and outcomes information is fed back to clinicians during treatment (Howard, Morass, Brill, et al., 1996; Sperry ,Brill, Howard, et al., 1996). However, the impact of this system on the quality of care, that is, clinical processes and outcomes, has not been determined. While many support the application of CQI in health care settings (Graham, 1995), there have been very few demonstrations of its implementation, much less its effectiveness.

Now it would be wonderful if we could say CQI is the magic bullet we are all waiting for. But it is no panacea. To say otherwise would fall prey to the misconception that for every complex problem there is a simple solution, which is usually wrong. The application of CQI to the mental health field, however, is not a naive translation of the approach used in industry. While we are proponents of CQI, we also think that we must be aware of the difficulties we face in applying this procedure to the mental health field. To implement a successful CQI program we need a long-term commitment. Otherwise, it will be just another human services fad. We want to describe some of these difficulties so we are all aware of the resources and time needed to show progress in this field.

Difficulties in Applying CQI to Child Mental Health Services

The simple statement that we need to supply information about process and outcome requires unpacking if we are to use CQI principles in mental health. It is clear that we need accurate and solid knowledge, but about what? First we need to be able describe the treatment. The heterogeneity of children's mental health services and practices makes it difficult to apply those CQI standards that prescribe as little variation as possible in producing goods or services (Dickens, 1994). The variability in service delivery is due to a variety of factors that are characteristic of mental health services. Among these are unreliability in diagnosis, multiple referring parties, different levels of parent and client adherence, and the different degrees of experience and expertise of the clinician. But most important is the dearth of standardized treatments. Unless the clinician follows a protocol or manual it is very difficult to describe mental health treatment. This difficulty is due to several aspects of therapeutic treatment. Mental health services are intangible and heterogeneous. The services represent processes that cannot be seen, or touched, and are difficult to articulate (Dickens, 1994).

Let's compare mental health treatment to an assembly line, since that is one of the examples typically used to illustrate CQI (Rust & Oliver, 1994). But we want to be very clear that we do not believe that mental health treatment should be like an assembly line. What we want to illustrate is that the transition from an approach developed primarily for manufacturing is not just simply applied to mental health. First, there is the difference in confidentiality in therapy that is not present in manufacturing. That is, the mere ability to observe is tempered by concerns about confidentiality and even if observation were possible would the presence of an observer affect the therapeutic process? Second, what a therapist does, that is his or her observable behavior, may not reflect the objectives of therapy. The manifest or observable aspects of treatment may only be indirectly related to what the therapist is trying to accomplish. Probably the biggest difference is that an assembly line is clearly laid out with specified operations. Mental health services are often not specified. Clinicians provide the treatment they deem appropriate. They have a great deal of autonomy in deciding what treatment and services should be provided. Unless the therapist follows a standardized protocol it is difficult to determine if the treatment is being delivered as planned, or even if there is a plan. Even under the latter conditions the therapist will only be guided in a general approach and not in a detailed schedule or plan. Finally, the assembly line is basically a closed system where there is very little or no impact of the environment on the product. In contrast, mental health treatment may play a small role in a person's complex environment. The assembly line has none of these characteristics. Again, we are not saying that mental health treatment should be like an assembly line. We are simply trying to point out the difficulty in apply some of the concepts of CQI to mental health treatment.

However, we must go beyond describing the treatment process. We must be able to focus on those aspects of the treatment process that lead to desired outcomes. Can we clearly identify those aspects of the client-therapist interactions that are critical to a successful outcome? The linkages between treatment processes and outcomes must be established before we can provide feedback on the success of those processes. At present we do not have the essential information that can guide us here. In the assembly line we can test each manufacturing process to determine if it is essential to producing the final product. There is another major difference we need to take into account. In the assembly line the product being manufactured has no relationship with the operator of the line. In therapy, the relationship between the consumer and the clinician is critical and the behavior of the client can affect the clinician. This interaction must be taken into account in applying CQI to mental health.

Third, we must be able to measure processes and outcomes in order to monitor them validly and reliably. Unless we have good measurement the feedback of information to the clinician will be misleading and possibly result in poorer outcomes. In the manufacturing analogy if we report erroneously that the outcome of the process is a product that weighs too much, when in fact it weighs too little, then reducing the weight of the ingredients will make the outcome even worse. Moreover, if our ability to change the weight, that is the process, is also unreliable then the errors can multiply. In mental health if we incorrectly inform the clinician that client is improving, then changes in treatment may be harmful. Thus, reliable and valid measures of key processes and outcomes are required for an effective CQI system

Fourth, we need to know how to return gathered information to those delivering the services. Can clinicians tell us what they consider important in managing services? We hoped that was the case. We surveyed 500 clinicians in United States and asked them to rate 21 different outcome and process measures (Bickman, Rosof, Salzer, et al., 1998). While there were distinct preferences, we found no systematic way to classify these preferences. Moreover, preferences did not seem to vary by discipline or experience. This suggests that clinicians do not have a clear conceptualization of the potential usefulness of measures in this area. In fact there is no good reason to expect that these measures should make sense to clinicians. Since systematic measures are rarely used in practice, clinicians do not usually have the experience necessary to have a conceptual grasp of the area. It appears that the simple approach, “just ask them”, is not going to work. Instead, we will need to conduct research to connect process measures to outcome data in order to identify which processes make a difference. In the assembly line analogy, the current situation is like giving feedback to the assembly line that is irrelevant to the manufacturing process.

Fifth, even if we are able to identify the critical processes and outcomes, measure them successfully in a clinical setting and provide the clinician with this potentially useful information we still need to be able to indicate how that information can be used. In our assembly line analogy if the product is too heavy, then the person controlling the assembly line will only find this information useful if there is a way to control the weight of the product. If this is as simple as dialing back on the amount of raw material and the operator knows how to accomplish this then we have an effective feedback loop. However, if the clinician is told that the therapeutic alliance score is dropping there are no proven methods that the clinician can apply to improve the alliance. By proven we mean that there is systematic evidence that if the clinician engages in certain behavior that alliance will improve. We do not believe that we have this type of information at this time.

Finally, even if we have built a better mousetrap, we have some doubt that the world of practice will beat a path to our door. There are both real and imagined concerns about the introduction of CQI measurement requirements. Human service organizations are difficult to change. Anything that limits the autonomy of providers is not well accepted by them. Some fear that confidentiality will be compromised. There are concerns that the data will be misused. Clinicians also worry that measurement will interfere with treatment. Thus, even a superior process may be difficult to implement. We do not place the full burden of resistance on the clinical community. In a field where are there insufficient resources to pursue business as usual, it is difficult to institute changes that require additional resources. Even cost effective procedures can have high initial cost. The change in organizational culture is probably the most difficult change that needs to occur. Our service organizations need to become learning organizations. They need to expand knowledge to improve services. The research community can not do this alone. Service organizations, government, consumers, and researchers must cooperate to achieve improved services. There will be costs and risks for all parties but the status quo is not acceptable

In summary there are several major barriers to the improvement of services that must be dealt with if we are going to achieve true progress. We need to have standardized treatments that have been shown to work in the real world. This will require more research on identifying effective treatments. Second we need to establish consensus on valid and reliable measurement of both process and outcomes. Accompanying the development of measures is the need to introduce them into the practice community. However, we must avoid a quick fix by using inexpensive measures that are not valid. This will require additional resources for hard pressed service organizations. With the proper resources and training, service organizations should be able to move into the forefront of this movement in partnership with others to create truly effective services for our children and youth.

References

Behar, L. (1988). An integrated state system of services for seriously disturbed children. In J. G. Looney (Ed.), *Chronic Mental Illness in Children and Adolescents* . Washington, DC: American Psychiatric Press.

Berwick, D. M. (1993). Sounding board: Continuous improvement as an ideal in health care. In A. F. Al-Assaf & J. A. Schmele (Eds.), *The textbook of total quality in health care* (pp. 31-39). Delray Beach, FL: St. Lucie Press.

Bickman, L. (1996). Special Issue: The Fort Bragg Experiment. *The Journal of Mental Health Administration, 23*.

Bickman, L. Guthrie, P. R., Foster, E. M., Lambert, E. W., Summerfelt, W. T., Breda, C. S., & Heflinger, C. A. 1995). *Evaluating Managed Mental Health Services: The Fort Bragg Experiment.* New York, NY: Plenum Press.

Bickman, L., Noser, K., & Summerfelt, W.T. (in press). Long term effects of a system of care on children and adolescents. *The Journal of Behavioral Health Services and Research*.

Bickman, L., & Peterson, K. A. (1990). Using program theory to describe and measure program quality. In L. Bickman (Ed.), *Advances in Program Theory*. (pp. 61-72). San Francisco: Jossey-Bass.

Bickman, L., Rosof, J., Salzer, M., Summerfelt, W.T., Noser, K., Wilson, S.J., & Karver, M.S. (1998) What information do Clinicians Value for Monitoring Adolescent Client Progress, unpublished manuscript.

Bickman, L., & Salzer, M., K. (1997). Special Issue: Measuring Quality in Mental Health Services. *Evaluation Review, 21*.

Borgesnstam, C., & Sandstrom, A. (1995). *Why the Vasa capsized*. Stockholm. Sweden: AB Grafisk Press.

Burns, B., Farmer, E., Angold, A., Costello, E. & Behar, L. (1996). A Randomized Trial of Case Management for Youths with Serious Emotional Disturbance. *Journal of Consulting and Clinical Psychology, 25(4)*, 476-486.

Deming, W. E. (1982). *Quality, productivity, and competitive position*. Cambridge, MA: MIT, Center for Advanced Engineering Study.

Deming, W. E. (1986). *Out of the crisis*. Cambridge, MA: MIT, Center for Advanced Engineering Study.

Dickens, P. (1994). *Quality and Excellence in Human Services*. Chichester: John Wiley & Sons.

Donabedian, A. (1986). Criteria and standards for quality assessment and monitoring. *Quality Review Bulletin*, 12, 99-108.

Dunne, J.E. (1997). Introduction: History and development of the practice parameters [Supplement]. *Journal of the American Academy of Child and Adolescent Psychiatry, 36(10)*, 1S-3S

Edwards, E. (1991). Total quality management in higher education. *Management Services, 35(12)*, 18-20.

Foster, E. M., & Bickman, L. (1998). *Refining the Cost-Effectiveness Analyses from the Final Report of the Outcome & Cost/Utilization Studies of the Fort Bragg Evaluation* . Nashville, TN: Vanderbilt University.

Fried, R.A. (1992). A Crisis in health care. *Quality progress, 24(4)*, 67-69.

Friedman, R., & Duchnowski, A.J. (1990). Children's Mental Health Services. *Journal of Mental Health Administration, 17(1)*, Special Issue.

Glazer, S. K. & Gaintner, R. (1995). Hospital administrators in an era of change and increasing responsibility. In L. E. Markson & D. B. Nash, (Eds.), *Accountability and Quality In Health Care: The New Responsibility*. Oakbrook, IL: The Joint Commission on Accreditation of Healthcare Organizations.

Gleick, J. (1992) Genus: *The Life and Science of Richard Feynman*, New York, Pantheon Books.

Graham, N. O. (1995). *Quality in Health Care, Theory, Application and Evolution*.

Gaithersburg, MD: Aspen Publishers, Inc.

Hoagwood, K., Hibbs, E., Brent, D., & Jensen, P. (1995). Efficacy and Effectiveness in Studies of Child and Adolescent Psycotherapy. *Journal of Clinical and Consulting Psychology, 63*, 683-687.

Howard, K. I., Moras, K., Brill, P.L., Martinovich, Z., & Lutz, W. (1996). The evaluation of psychotherapy: Efficacy, effectiveness, patient progress. *American Psychologist, 51*, 1059-1064.

Ivancevitch, D.M., & Ivancevitch, S.H. (1992). TQM in the classroom. *Management Accounting, 74(4)*, 14-15.

Laffel, G., & Blumenthal, D. (1989). The case for using industrial quality management science in health care organizations. *Journal of the American Medical Association, 262*, 2869-2873.

Markson, L.E.& Nash, D.B., (1995.). *Accountability and Quality in Health Care: The New Responsibility*. Oakbrook Terrace, IL: The Joint Commission on Accreditation of Healthcare Organizations.

McGlynn, E.A., Halfon, N. & Leibowitz, A. (1995). Assessing the quality of care for children: Prospects under health reform. *Archives of Pediatric and Adolescent Medicine, 149*, 359-368.

Nash, D. (1995, October). *Quality and utilization management: The cutting edge for 2000*. Presentation at the American Managed Care and Review Associations Quality and Utilization Management Conference, Nashville, TN.

Nathan, P.E. (1998). Practice guidelines: Not yet ideal. *American Psychologist, 53(3),* 290-299.

Oakland, J. S. (1989). *Total Quality Management, Oxford: Heinemann Professional Publications.*

O'Brien, J. (1990). Developing high quality services for people with developmental disabilities. In V.J. Bradley & H.A. Bersani, (Eds.), *Quality Assurance for Individuals with Developmental Disabilities: It's Everybody's Business. Baltimore, MD: Paul H. Brooks Publishing*.

Panzer, R. J. (1995). Are health care quality improvement and public accountability compatible? In L. E. Markson & D. B. Nash, (Eds.), *Accountability and Quality in Health Care: The New Responsibility*. Oakbrook Terrace, IL: The Joint Commission on Accreditation of Healthcare Organizations.

Rivera, V. R., & Kutash, K. (1996). *What works in children's mental health services? : Uncovering answers to critical questions*. Paul H. Brookes Pub. Co., Baltimore, 1996.

Rust, R. T., & Oliver, R. L. (1994). Service quality: Insights and managerial implications from the frontier. In R.T. Rust & R. L. Oliver, (Eds.), *Service Quality: New Directions in Theory and Practice*. Thousand Oaks, CA: Sage.

Salzer, M. S., Bickman, L., & Lambert, E.W. (1999). Dose-effect relationship in children's psychotherapy services. *Journal of Clinical and Consulting Psychology, 66(2),* 270-279.

Sawyer. M., Sarria, A., & Baghurst,P. (1991). The Use of a Computer-assisted Interview to Administer the Child Behavior Checklist in a Child Psychiatry Service. *J.Am.Acad.Child Adolesc.Psychiatry, 30*, 674-681.

Saxe, L., Cross, T., Silverman, N., Batchelor, W. F., & Dougherty, D. (1987). *Children's mental health: Problems and treatment*. Durham, NC: Duke University Press.

Smith, H. (1988). *Collaboration for Change: Partnership between Service User's, Planners and Managers of Mental Health Services*. London: King's Fund Centre.

Smith, M. L. & Glass, G.V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist, 32*, 752-760.

Sperry, L., Brill, P., Howard, K.I., & Grissom, G. (1996). *Treatment outcomes in psychotherapy and psychiatric interventions*. New York: Brunner/Mazel.

Stroul, B. A., & Friedman, R. (1986). *A system of care for children and youth with severe emotional disturbances*. Washington, DC: Georgetown University Child Development Center, CASSP Technical Assistance Center.

Swiss, J.E. (1992). Adapting total quality management to government. *Public Administration Review, 52(4)*, 356-362.

Weisz, J. R., Weiss, B., Alicke, M.D. & Klotz, M.L. (1987). Effectiveness of Psychotherapy with Children and Adolescents: A meta-analysis for Clinicians. *Journal of Consulting and Clinical Psychology, 55*, 542-549.

Weisz, J. R., Weiss, B., Han. S.S., Granger, D.A., & Morton, T. (1995). Effects of psychotherapy with children and adolescents: A meta-analysis of treatment outcome studies. *Psychological Bulletin, 117*, 450-468.

Zieman, G. L. (1995). *The complete capitation handbook: How to design and implement at-risk contracts for behavioral healthcare*. Tiburon, CA: CentraLink Publications.

APPENDIX 8

THE IMPLEMENTATION OF A CLINICIAN-BASED ROUTINE OUTCOME MEASURE (HONOSCA) IN A CHILD AND ADOLESCENT MENTAL HEALTH SERVICE

Peter Brann
Lecturer in Psychology
Maroondah Hospital CAMHS
Monash University Departments of Psychological Medicine and Psychology
E-mail: pbrann@silas.cc.monash.edu.au

In the construction of a comprehensive outcome measurement system, it is important that the perspectives of all parties are included. Children, adolescents, carers and clinicians each have different domains of information and different biases. Clinicians have access to a range of knowledge about mental health states. However, relatively few instruments are suitable for routine use by clinicians (Hunter, Higginson and Garralda, 1996). The Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA) is a promising, though very new, instrument that could potentially balance the competing demands of practicality and psychometric rigour.

HoNOSCA comprises thirteen core scales (rated between 0 to 4) and two optional scales. HoNOSCA ratings are based on whatever sources of information the clinician has available. The scales address behaviours, symptomatology, disability, and social functioning (Gowers, Whitton, Harrington, Beevor, Lelliot, Wing, Curtis and Jezzard, 1999; Gowers, Harrington, Whitton, Lelliot, Beevor, Wing and Jezzard 1999). Specifically the scales address the following:

- Disruptive/Aggressive Behaviours
- Overactivity/Concentration
- Self-injury
- Substance misuse
- Scholastic/Language Skills
- Illness/Disability

- Hallucinations/Delusions
- Non-organic Somatic Symptoms
- Emotional Symptoms
- Peer Relationships
- Self-care
- Family Relationships
- School Attendance

The Maroondah Hospital Child Adolescent Mental Health Service has implemented HoNOSCA as part of a strategy to develop an outcome-informed system of care. The Maroondah service covers a large suburban and semi-rural part of Melbourne and has the modest resource levels characteristic of CAMHS. All administrative processes and training demands were designed to minimise the impact on a busy public health service. Clinicians were asked to complete HoNOSCA at initial assessment, 3 months, 6 months (and subsequent 6-month periods) and at discharge. With the exception of the 3 month mark (which was later removed), all times corresponded with existing clinical review periods.

Following approximately two hours of training, clinical vignettes, (developed by staff blind to HoNOSCA) were rated by all clinical staff. This resulted in reliability estimates of good to very good for the majority of the scales. Reliability for the total scale was good. The reliability of an instrument varies according to the purposes for which it is to be used. As with many other instruments, HoNOSCA had higher reliability when estimating the consistency of a clinician's responses compared with consistency of estimating absolute scores. In other words, using HoNOSCA to examine aggregated results within a service or a caseload is a more reliable venture than using HoNOSCA to compare individual clinicians or establish a cut-off for "caseness".

Our initial analyses were based on 600 HoNOSCA records. Substantial data collection and analysis have occurred over the past 2 years and are continuing. Psychometrically, the instrument was found to have a coherent internal structure. Factor analysis extracted four factors, which were found to match the clinical clusters of externalising, psychotic personality, learning disability and internalising disorders. The total score significantly correlated with higher scores on the subscales indicating support for the notion of the total score as a proxy for severity. Scale scores were related to clinical diagnosis.

Differences in age and gender scores on the 13 scales were consistent with clinical expectations. When clinicians were blind to their first HoNOSCA score, their subsequent HoNOSCA scores were significantly related to their retrospective global impressions of change.

Over a three-month period, HoNOSCA was sensitive to change with 10 of the 13 scales showing improvement. Although not significant, a trend towards worse ratings on substance use at time 2 was seen as congruent with clinical impressions that this information was rarely fully disclosed at assessment. Although based on a small sub-sample, and currently being re-examined, HoNOSCA change scores revealed a strong trend toward a difference between those discharged at time 2 and those still in ongoing care. Further study in two different hospitals and day program settings shows higher HoNOSCA scores in those locations compared with clients in outpatient settings suggesting those with more severe presentations are being treated in these more resource intensive service locations.

Clinicians reported that HoNOSCA was brief to complete and accesses many, though not all, of the mental health domains on which children could change. Interestingly, clinicians have used HoNOSCA results in a variety of imaginative ways, for example: prioritising cases for review in supervision; highlighting cases with deterioration; self-reflection; structuring thinking about case progress; and as an infrastructure for evaluating groups and programs. A number of clinicians have pioneered sharing graphs of HoNOSCA results with families as a means of enhancing collaboration and transparency. Details of these analyses can be found in Brann (1999).

Current studies are investigating further the properties of the instrument, the utility and impact of providing graphical feedback, and the relationship of the instrument to carer and client measures. Maroondah operates from the orientation of a learning organisation (Birleson, 1998) and values the provision of an outcomes infrastructure and feedback loops to facilitate individual and service learning. We decided not to prescribe how HoNOSCA must be used and believe that this has helped contribute to clinicians valuing of outcome measurement and HoNOSCA.

While Maroondah Hospital CAMHS makes no definitive claims as to the long-term standing of this initiative, measures such as HoNOSCA and the processes utilised by us do appear to allow clinicians and services to monitor and learn from outcomes. HoNOSCA does appear to be suitable, pragmatically and psychometrically, for routine clinical use. In the movement towards comprehensive outcome measurement, it allows for clinicians’ perspectives to be included.

References

Birleson, P. (1998) Learning organisations: A suitable model for improving mental health services? *Australian and New Zealand Journal of Psychiatry, 32*, 214-222.

Brann, P. (1999) Outcome Measures and Quality Improvement in CAMHS, Maroondah Hospital CAMHS/Monash University Departments of Psychological Medicine and Psychology.

Gowers, S., Harrington, R., Whitton, A., Lelliot, P., Beevor, A., Wing, J. and Jezzard, R. (1999) Brief Scale for measuring the outcomes of emotional and behavioural disorders in children. *British Journal of Psychiatry, 174*, 413-416.

Gowers, S., Whitton, A. and Harrington, R., Beevor, A., Lelliot, P., Wing, J., Curtis, R. and Jezzard, R. (1999) *HoNOSCA Health of the Nation Outcome Scales for Children and Adolescents Report on Research and Development*, University of Liverpool, September 1995-March 1997.

Hunter, J., Higginson, I., & Garralda, E. (1996) Systematic literature review: outcome measures for child and adolescent mental health services, *Journal of Public Health Medicine, 18(2)*, 197-206.