

David C. Factor, Ph.D., C.Psych.
phone/fax (416) 631-7861

Robert J. Fulton, M.S.W.
phone (416) 481-7803
fax (416) 481-4299

PRO.FILES
Associates in Research
and Evaluation
97 Chudleigh Ave
Toronto, Ont. M4R 1T4
r.fulton@home.com

Research Methodology Manual

Risk Screening and Outcome Evaluation for OARTY

This document is part of the OARTY Information System (OIS) © copyright 1999, all rights reserved

Table of Contents

1.0	Description of the OARTY risk screening system:	2
1.1	Values and Functions of the system	2
1.11	risk screening.....	2
1.12	Global Outcome Data.....	3
1.13	Support Front Line staff	3
1.14	Oriented to Strengths.....	3
1.15	Accountable to Children and Parents	4
1.16	Accountable to Principles of Ethical Research	4
1.17	Good coverage of the issues with the Lowest Administration Costs	4
1.18	Management systems to ensure standard of excellence	5
1.2	The Criteria for Measuring Risk	6
1.21	Health and Safety is measured by:	6
1.22	The child's or the family's ability to cope is measured by:.....	7
1.3	Definition of Risk Classifications	7
1.4	Outcome Evaluation.....	8
1.41	Changes in the Scores on the Risk Screening Instruments.....	8
1.42	Defining and Measuring Standardized Service Results	8
1.5	Plan of care.....	9
1.51	Assessment of the child's needs and reasons for service	10
1.52	Priorities of the Plan of Care	10
1.52a	Barriers	10
1.52b	strengths	10
1.52c	goals	10
1.52d	service diary	10
1.6	Reviewing the Results of Service.....	11
1.7	Measuring the Results of Service.....	12
2.0	The goals of this manual are:	13
3.0	the Logical and Empirical Basis of Risk Screening	13
4.0	The Multigate Screening Structure	15
5.0	Measures of Health and Safety	17
5.1	Conners' Global Index	17
5.2	Children's Global Assessment Scale (CGAS)	19
5.3	The Symptom Assessment - 45 (SA-45).....	20
5.4	The Feelings, Attitudes and Behaviours Scale for Children (FAB-C).....	21
5.41	Relationship of the FAB-C to other Self-report measures for children...	21
5.42	Reliability and Validity of the FAB-C	22
5.5	Children's Autistic Rating Scale	23
5.6	The Level of Assistance Scale (LA).....	23
	Measuring Risk with the Level of Assistance Scale	24
	Validity of the LA	25
5.7	Conclusion on Premorbid Risk Factors.....	26

6.0	Measures of the Child’s & Parent’s Ability to Cope	26
6.1	the Objective Stressors Checklist.....	26
6.2	the Questionnaire on Resources and Stressors (QRS-F).....	28
6.3	the Socio demographic checklist.....	29
6.31	How is the sociodemographic checklist of risk factors different from the objective stressors checklist?	31
6.4	the Parental Bonding Instrument (PBI).....	32
6.5	the Vineland Adaptive Behaviour Scale (VABS).....	34
6.6	Conclusion on Measures of the Child’s or Family’s Ability to Cope.....	34
6.7	The Full Set of Screening Instruments.....	35
6.71	Universal instruments.....	35
6.71	medically fragile.....	36
6.72	autistic.....	36
6.73	dual diagnosis (DH + conduct or psychiatric disorder).....	36
6.74	conduct and emotional disorders.....	36
6.75	socially disadvantaged, abused and under-developed.....	36
7.0	Outcome Measurement	37
7.1	Indicators of Success.....	38
7.2	second level - individualised outcome measurement.....	38
7.21	the Social Skills Checklist (Gresham and Elliot).....	38
7.22	the WRAT-R and the K-TEA.....	39
7.23	the Interpersonal Support Evaluation List (ISEL).....	39
7.24	The Coping Inventory for Stressful Situations (CISS).....	39
7.3	the Plan of Care Model.....	40
7.31	Reactivity.....	41
7.32	Drift.....	41
7.33	Expectancies.....	42
7.34	Lying.....	42
8.0	Item Analysis of the Indicators of Success	42
8.1	Vision #1 Basic Functioning.....	42
8.2	Vision #2 Social Support.....	43
8.3	Vision #3 Safe, Supportive and Accepting Care Givers.....	43
8.4	Vision #4 Health promotion and avoiding risk taking behavior.....	45
8.5	Vision #5 Life Skills.....	47
8.6	Vision #6 Resiliency (ability to cope).....	47
8.7	Vision #7 success at school or work.....	48
9.0	Threats to the Truth in Outcome Research	49
9.1	regression from the mean.....	49
9.2	test data is not a substitute for real life outcomes.....	49
9.3	all tests have a built in region of false positive and false negative.....	49
9.4	highly sensitive tests are vulnerable to short term, transitory events -.....	50
9.5	alpha inflation.....	50
9.6	sleeper effect.....	50
9.7	The examiner test taker relationship.....	51

9.8	Expectancy effects.....	51
9.9	Effects of reinforcing responses.....	51
9.10	The absence of control or comparison groups.....	52
9.11	ensuring a truly representative clinical group.....	52
9.12	the lack of random assignment.....	52
9.13	anchoring bias	53
9.14	available heuristic bias	53
9.15	covariance heuristics - bias.....	53
10.0	Two Fundamental Designs to identify True Outcomes	54
10.1	the single case design	54
10.2	the Group design	55
10.3	accountability for the single subject.....	56
11.0	Hot Issues in the Practice of Testing	57
11.1	Professional Issues	57
11.2	Moral Issues	57
11.3	Ethical Issues.....	58
11.4	Social Issues	59
	Bibliography	61

Research Methodology Manual

Risk Screening and Outcome Evaluation for OARTY

synopsis:

This manual is designed to document each component of the risk screening and outcome evaluation methodology, including the selection of tests, the identification and control of errors in the administration, coding and interpretation of data, and the empirical basis for the content of the system.

OARTY Mission Statement:

- To develop and coordinate resources to benefit children and youth in need who are physically or sexually abused, emotionally disturbed, developmentally challenged, autistic, medically fragile, young offenders, dually diagnosed, conduct disordered, psychological and psychiatric disorders, and other hard to serve children living in residential care.
- To provide support, encouragement and to facilitate the healthy development of these children and youth.
- To liaise and consult with other children's and youth's service providers to maximize efforts and initiatives.
- To create, maintain and share relevant communications among service providers, internal and external to the Association.

What does OARTY do?

- OARTY is a provincial network of children's residential services. We represent over 70 member agencies and over 2000 residential beds for children across Ontario. We are licensed and regulated under the Child and Family Services Act and funded on a per diem basis.
- OARTY provides residential care and treatment for children and youth who are physically and sexually abused, emotionally disturbed, developmentally handicapped, autistic, medically fragile, young offenders, dual diagnosed, conduct disordered, psychiatric and psychological disordered, and other hard to serve children and youth. Our services include residential care, specialized foster care, treatment facilities, education and day treatment.
- OARTY agencies are committed to treatment. We ensure the health, safety and welfare of children and youth in our care.

1.0 Description of the OARTY risk screening system:

1.1 Values and Functions of the system

The risk screening and outcome evaluation methodology is a mix of something old and something new. Agencies which provide residential care have been always been obliged under the terms of their license as set out in Ontario Regulation 70, section 86 (2) and section 111 to assess the needs of the children placed, to set goals, work out a treatment plan and review the results. The screening and evaluation process is not very different from what agencies have done for years. The changes made to improve the typical process in OARTY agencies are: {1} we are using a common language, especially to describe goals and treatment plans; {2} the data is being collected into a computerized database and {3} a small number of standardized risk screening instruments are being used by everyone.

The commitment to improve the typical assessment/case planning system was inspired by a Government policy document (*Making Services Work for People, a new framework for children and people with developmental disabilities*). The goals and future directions of the Ministry's *Making Services Work* policy framework, include:

goal #2	“those most in need will receive essential supports”
goal #5	“families and individuals will receive services that respond to their needs” ¹
future direction #1	“Moving to outcomes-based funding”

In addition, improvement made to the assessment/case planning process incorporated a number of ethical principles. Other criteria for upgrading the system included low administrative costs (in terms of worker time and the price of the tests).

The functions of the upgraded assessment/case planning system are as follows:

1.11 risk screening

The Ministry's “Goal #2” (*those most in need will receive essential supports*) requires clear definitions and measurable criteria for identifying who is most in need, less in need and least in need. The fact that a child is in the care of the CAS means that he is “most in need” compared to

¹The text of goal #5 states:

“Ensuring services are provided that can respond to user's needs means designing programs that are flexible, adaptable to change and able to build upon existing strengths. In many cases, this will permit individuals and families to exercise greater control and direction over the kinds of services they receive”, page 6, *Making Services ...*

the general population. However, we still require a method to measure the needs of children in care in order to allocate intensive interventions. OARTY defined “need” in terms of the domains that matter most to public policy, the risk for bad life span outcomes.

Function **1.11** *to have clear definitions, common decision criteria and standardised instruments for classifying children in relation to high, medium and low risk for bad life span outcomes.*

The risk of bad life span outcomes refers to the probability that the child *is currently or will*: {a} behave in ways that are a danger to self or the public; {b} suffer from a serious debilitating mental illness; {c} commit crimes and be incarcerated; and {d} place their children in danger of neglect, sexual abuse and physical abuse. We assume that the intent of provincial policy is to allocate scarce intensive services to those who are most likely to exhibit very bad life span outcomes without effective intervention.

1.12 Global Outcome Data

The Ministry’s “future direction #1” (*Moving to outcomes-based funding*) assumes that we have a system in place for measuring the changes in client functioning over the domains that matter most to public policy (and related to the definition of “most in need”). The term global outcome data refers to a standard test protocol used for all clients receiving OARTY services.

Function **1.12** *by re-testing the child with the risk screening instruments, to keep track of the risk level as the child moves through service (in this sense, the instruments serve as global outcome data)*

The Ministry’s “Goal #5” (*families and individuals will receive services that respond to their needs*) assumes that we have a system that {a} supports a high standard of assessments and plans of care; {b} identifies the strengths of children and families; {c} involves children and families directly in identifying the goals of service; and {d} gives the children and families full information on the results achieved for both risk screening and outcomes. This aspect of provincial policy led to additional functionality in the OARTY system:

1.13 Support Front Line staff

Function **1.13** *to support front line staff (care givers, child care workers and social workers) to make better assessments and better plans of care*

1.14 Oriented to Strengths

Function **1.14** *to collect data on the child’s strengths, skills and accomplishments to define the goals of the plan of care in terms of skills and accomplishments, rather than in terms of negatives or pathology*

1.15 Accountable to Children and Parents

Function 1.15 *to measure whether the goals set in the plan of care are achieved in the opinion of the child himself, parents, the CAS worker, teachers and community partners*

1.16 Accountable to Principles of Ethical Research

We were guided by the following ethical considerations:

- Ethics [1] the language and content of standard screening instruments must not damage the trust between the child and the worker; they must not open up deep emotional issues that the service cannot deal with
- Ethics [2] the tests must not be diagnostic or administered as if they were diagnostic. Only medical doctors or psychologists are legally permitted to diagnose anyone.
- Ethics [3] the screening instruments will have known sociometric properties with acceptable levels of reliability and validity.
- Ethics [4] the danger of labelling children must be counter balanced by giving workers the power and training to over-ride the classification of risk given the probability of false positive and false negative errors.
- Ethics [5] Children and parents will be told on every occasion that the tests are administered, that {a} they do not have to answer the questions, {b} if they start answering, they may change their minds at any point and stop completing the test and [c] if they choose not to co-operate with the test protocols, this decision will not affect their care in any way whatsoever.

1.17 Good coverage of the issues with the Lowest Administration Costs

Some of the practical criteria that assures a cost effective solution includes:

- *keep it simple*: no test should be longer than one page
- *maximise the use of tests which are in the public domain*
- *empower the front line*: make sure that front line practitioners are both qualified and able to score and interpret the test themselves (*instantly*) without the aid of a psychologist or a computer

- *test results flow right into plan of care:* Families, children and front line workers must be able to respond to the data flowing from the outcome measurement. By ensuring the relevance and benefits of the system flow to the front line, we are better able to keep the front line committed to the data collection effort.
- *tests must cover a broad range of issues:* Don't put all your eggs in one basket. Cover behaviour, symptoms, stressors, adaptive functioning and social support.
- *Don't be dependent on only one strategy of testing.* Even the "best tests" have enormous error rates.
- *Don't be redundant:* The individual scales or scores coming off the system **must not be highly correlated.** For example, eighty percent of the children that score in the clinical range on aggression or delinquency will also be in the clinical range on the externalising scale of the Achenbach. (Achenbach, 1991) Because of this, you can't use both the aggression scale and the externalising scale, without being redundant. *This means that the risk screening and outcome criteria can use only one or two numbers coming off any one test.*

1.18 Management systems to ensure standard of excellence

Front line workers and foster parents administer, score and interpret the tests. The risk screening system collects between 4 and 6 numbers for each child at each risk screening cycle described below. The integrity of the test protocol is ensured through the following management systems:

- {1} Front line workers report the test results and their interpretation to their supervisor. The supervisors are responsible for ensuring that workers are administering the test properly and are in compliance with the ethical standards.
- {2} In large agencies, the work of the supervisors is reviewed by a senior member of the agency (such as the Clinical Director or a consulting psychologist). In small agencies this middle step {2} is eliminated.
- {3} The supervisors or the senior agency staff responsible meets regularly with the zone research co-ordinators to review issues arising in the test administration, scoring and interpretation.

- {4} The zone research co-ordinators are trained and qualified to administer, score and interpret the tests; moreover, they have a broad understanding of the limits of knowledge inherent in these tests, including their reliability and validity, the make up of the normative samples and the statistical methods which are at the core of the interpretation schemas.
- {5} The zone research co-ordinators have direct and immediate access to the research consultant, Robert Fulton. Journals, chapters and manuals relevant to the research are collected and disseminated to the zone research co-ordinators.
- {6} Three manuals have been written to support all levels of staff through the process:
 - {a} administering, scoring and interpreting the risk screening instruments;
 - {b} interpreting the outcome data including the goals and plan of care; and
 - {c} research methodology
- {7} A copy of the actual paper and pen test will be retained on the child's file in the home, along with the scoring guide and any notes taken by the person administering or scoring the test.
- {8} Once a month, the risk screening data and the goal assessment information from the OARTY Information System is transmitted to the research consultant, where it is assembled into an OARTY database under SPSS. Statistical reports flow from the system routinely and back into the agency management structure.

1.2 The Criteria for Measuring Risk

The system is based on “multi-gate” risk screening, in which children (or their families) are assessed on two primary domains:

- 1.21 health and safety of the child
- 1.22 the child's or the family's ability to cope

1.21 Health and Safety is measured by:

Children's Global Assessment Scale (CGAS)child's social impairment
 Conners' Global Index child's behaviour problems
 SA-45 teenager's psychiatric symptomatology
 or FAB-C school-age child's emotional symptomatology
 The Childhood Autism Rating Scale (CARS)..... degree of autistic symptomatology
 Level of Assistancedegree of care required for medically fragile children

The central belief of the tests at threshold {a} is that a child's health and safety is compromised by one of four, fairly distinct and uncorrelated conditions: (1) chronic behaviour problems, (2) an inability to function at home, in school or in the community, (3) the presence of high levels of anxiety, worries, depression and crazy thoughts, (4) evidence of autistic symptoms and (5) the presence of medical & or physical conditions requiring direct intervention for the child to function at the most basic level.

1.22 The child's or the family's ability to cope is measured by:

the child objective stressors checklist the child's experience of stress
 the QRS-F..... the parent's experience of stress
 a sociodemographic checklist.....adverse family history
 parental bonding instrument..... quality of attachment
 Vineland Adaptive Behaviour Scale (VABS)..... child's developmental functioning

The central belief behind the second threshold is that the *probability that the child will get worse without intervention* is determined by a combination of

1. The child's *current stress levels*
2. The parent's *current stress levels*
3. The *accumulation* of very stressful events and adverse conditions from early life
4. The quality of the child's *attachment* to his or her primary care givers
5. The child *acquisition of basic skills* in communication, socialisation, daily living and motor development compared his peers

1.3 Definition of Risk Classifications

No child will be given all of the tests under each threshold. At a minimum, a child will be given one test per threshold (the CGAS and the socio-demographic checklist). In most cases, a child will be given two tests under each threshold and the highest score (or worst score) is the result used in the classification of risk, according to the following rules:

- very high risk* (if both thresholds are in the "high risk" zone)
- high risk* (combination of high and medium across the two thresholds)
- medium risk* (both thresholds are in the medium zone or a combination of high-low across the two thresholds)
- low risk* (both thresholds are in the low risk zone or a combination of medium-low across the two thresholds)

Each test has three “zones”: a high risk zone (e.g. a T-score of 80 or greater on the Conners’ Global Index); a medium risk zone (e.g. a T-score between 66 and 79 on the Conners’ Global Index); and a low risk zone (e.g. a T-score of 65 or lower on the Conners’ Global Index). Please see the Summary Sheet for Determining the Risk level in the appendix for complete details on the “risk zones” of each test.

1.4 Outcome Evaluation

The outcomes for the child or family are measured in two ways:

1.41 by changes in the individual scores on the tests of the risk screening system

The risk screening tests will be administered once every nine months and at discharge from the resource (if this occurs outside the nine-month cycle)

1.42 by observing changes within a standard plan of care model incorporating the child’s priorities in terms of learning skills and accomplishing personal milestones

1.41 Changes in the Scores on the Risk Screening Instruments

By re-testing the child with the risk screening instruments, the instruments serve as global outcome data. The term global outcome data refers to a standard test protocol used for all clients receiving OARTY services.

1.42 Defining and Measuring Standardized Service Results

On admission and every nine months thereafter (and at discharge), the OARTY worker, the child and his parents together complete a checklist entitled “Indicators of Success”. The checklist is made up of seven vision statements, life span outcomes related to each vision and “indicators of success”. The indicators of success are the standard service results of OARTY agencies. The interventions of OARTY agencies are directed towards helping children and families to be successful *on these terms*. For any individual child, the interventions will only work towards a limited number of these results, depending on what are the priorities for child or his family. Children and parents assess the status of all possible goals in relation to four categories:

S or strength - the child functions very effectively

OK the child has no problems or frustrations functioning

NW or needs work - the child has not had a chance to learn the skills or is frequently frustrated while attempting it; but this area is reserved for future work

PR should be a priority for the upcoming plan of care

The indicators of success were taken from the literature according to the following criteria:

- (1) the content of the indicators are based on the skills and accomplishments for the child and family - every item is defined in positive terms
- (2) the specific skills and accomplishments are those which predict the best life span outcomes for the child

The validity of the assessment of the child's success on all indicators is supported in the following ways:

- {1} The assessment of the child's success on all indicators is discussed in the plan of care conference and signed off by the CAS worker, group home staff, the parents and the child
- {2} The indicators most frequently selected for the plan of care (marked "PR" by the child and his parents) will have a program of complementary testing, which measures the degree of skill acquisition. The complementary tests include the *Social Skills Checklist*, the *ISEL* and the *CISS*. Other tests may be used as appropriate.
- {3} Most of the indicators are assessed through direct observation, since the descriptions of the indicators are concrete and observable.
- {4} With every indicator selected for the plan of care, *the plan of care model* ensures that conclusions about the results of service are reasonable and accountable.

1.5 Plan of care

Every agency is required, as a condition of its license, to prepare a plan of care within 30 days of admission, to monitor the plan monthly and to review it completely every three months. Moreover, the child, his parents, CAS worker and community partners (such as teachers) *must be invited* to participate in the initial plan and in the quarterly reviews. The plan must be recorded and copies sent to all parties.

The full plan of care includes details on family visiting, the child's education, child welfare court, medical and dental care, important events in the child's life, the child's behavior problems and the consequences and actions required to meet any other need of the child or his family. In addition to these details, the plan of care includes an individualized treatment planning component, which is made up of the following parts:

1.51 Assessment of the child’s needs and reasons for service

This section includes the risk screening data, other assessment information, and the “Indicators of success” form, the actual scores and the clinical interpretation of the results of testing completed on or about the child.

1.52 Priorities of the Plan of Care

These are drawn from the instrument in section 1.42, *Defining and Measuring Service Results*. One of the classifications for each skill or accomplishment itemized is “a priority for the plan of care”. Each priority is discussed separately according to the following structure:

1.52a Barriers

The barriers are the forces that have prevented the child from learning the skill or accomplishing personal milestones - such as getting and keeping a job. If these forces are still active, they will have to be countered through the intervention in order for the child to grow. Examples of the common barriers facing children are: {a} a lack of opportunity for jobs or friendship {b} poor attitudes such as hopelessness or resentment, {c} lacking a “model” or a coach to teach them communication or social skills and {d} interpersonal issues, such as being too dependent on someone.

1.52b strengths

Strengths are the skills and goals that have already been achieved by the child; other strengths include special opportunities - such as a potential job placement or a new recreational program - that may provide an occasion and motivation to learn a skill.

1.52c goals

The goals are the stages that a child goes through in order to be successful in accomplishing a personal milestone or learn a skill. The goals are conceptualized by breaking down the “learning task” into small, fully achievable steps for the child to take. The goals should be defined in clear behavioral terms; it should be easy for the child and everyone to know if the child was succeeding in terms of the goals. The goals will be unique for each child, since every child will have a different pathway to their learning objective, depending on their stage of development, barriers and strengths.

1.52d service diary

The plan of care includes a list of tasks and actions that everyone agrees to do, including details on who will do what and when. The record of who actually does what and when is

the diary of interventions and services provided. By keeping exquisite records of the service diary, it is possible to exactly measure the time expended and the cost for each result of service. OARTY will institute a computer based system to keep exquisite records of the time and dollar cost associated with obtaining the priority results for a particular child - but only for limited periods of time on certain cases.

The plan of care model provides a structure to record the results of service within the context of:

- {1} the child's development (in terms of the skills he has learned and his accomplishments to date)
- {2} the child's strengths (in terms of opportunities, social support and personal characteristics)
- {3} the child's challenges or barriers to overcome
- {4} the service inputs (in terms of staff, time & resources needed to support the child)
- {5} the goals that the child achieved in order to be considered successful (by his staff, parents and himself). These goals are stated in behavioural terms as observable, small steps on a pathway leading to the acquisition of a skill or the accomplishment of a personal milestone.

1.6 Reviewing the Results of Service

The plan of care model is a continuous process that is reviewed monthly for as long as the child is receiving service. During the monthly reviews:

- {1} The child's progress on his goals is assessed (as achieved or not)
- {2} The service inputs are recorded (in relation to the commitments in staff, time and other resources, to support the child)
- {3} The service diary and goals for the child are adjusted in response to new challenges or opportunities that may have appeared since the formal plan was agreed to

Every 90 days, the plan of care is reviewed in more complete detail. As part of the 90 day review, the participants to the original plan are invited to attend (the child, parents, foster parents, school teachers, CAS worker, etc.). At the 90 day review:

- {1} one or more priorities from the previous plan of care may be removed as "fully

achieved” if, in the judgement of the participants, the child has achieved his goals

- {2} one or more priorities from the previous plan of care may be removed as “partially achieved” if, in the judgement of the participants, the child has achieved some of his goals relative to the priority area
- {3} one or more priorities from the previous plan of care may be continued in to the next phase of the plan of care
- {4} new priorities may be added

During the 90 day review, all components of plan of care model (section 1.52) are reviewed and for each priority, the full cycle of information (section 1.52a..d) is updated and recorded.

Finally, there is a 9 month review of the plan of care. The 9 month review is different from the 90 day review in two respects:

- {1} the “Indicators of Success” are assessed once again with the child and his parents (or foster parents)
- {2} the child and his family are re-assessed on the same risk screening instruments used at the beginning of treatment

1.7 Measuring the Results of Service

The results of service are measured in strictly positive terms:

The personal skills and accomplishments of the child

Every nine months, the indicators of success are re-assessed. Any change in the classification of an indicator (in terms of strength, OK, needs work and priority). By using the statistical procedures associated with “survival analysis”, changes in each assessment of the plan of care can be analyzed in relation to “how much and in what way the child has changed”.

At the second level, the details of the plan of care model reveal which goals that the child achieved in order to learn a skill or accomplish a personal milestone. The information about goals achieved is anchored in a developmental context, in “time” and in the context of service inputs.

The plan of care model provides a way to link the child’s progress in skills and accomplishments on his personal milestones to the intervention itself. By controlling for the strengths and barriers, related to each priority area, we can isolate the main effects of intervention from the interaction effects of strengths and barriers in the child’s social context. Moreover, by controlling for the

“risk status” of the child, we can ensure that the comparison groups are equivalent. The only element remaining which explains the progress or lack of progress of the child is the service diary.

Finally, the global outcome data (section 1.41) provides a criterion against which inferences about the main effects of intervention can be validated. Indeed, the criterion is completely independent of the logic of the plan of care and the instruments (which constitute the criterion) are standardized, norm referenced, reliable and valid.

2.0 The goals of this manual are:

- 2.1 to document and explain the rationale for each decision in the methodology
- 2.2 to review the literature on each component of the research design
- 2.3 to examine the selection of tests, specifically, what options were considered, why each test was chosen and why other options were not chosen
- 2.4 what are the threats to the truth on both the clinical & outcome evaluation levels
- 2.5 how do we minimize these threats
- 2.6 what are the professional, ethical and moral obligations of the research

3.0 the Logical and Empirical Basis of Risk Screening

Because the primary purpose of the tools is to screen cases for the risk of bad life span outcomes and to allocate scarce intensive interventions, the concept of *risk* is much more restrictive and severe than is implied in prevention programs or epidemiological research. (Rutter & Sandberg, 1985) Rutter and Sandberg (1985) point out that children who require intensive treatment for truly debilitating disorders are very different than children who present symptoms and behavior problems in broadly based epidemiological surveys. Rutter and Sandberg (1985) state that children who are at risk for developing enduring serious mental health problems are recognizable through the application of four criteria:

- {a} Certain behavioral symptoms are much more predictive of bad outcomes than others. The two strongest predictors are poor peer relationships (measured by the FAB-C) and hyperactivity/inattention (measured by the Conners' Global Index)
- {b} Disorders that are more pervasive over situations are more persistent over time. In other words moderate problems in school, at home and in the community is more serious (risky) than even more serious problems in one place, such as school. This fact is picked up by the instrument, the CGAS. (See section 1.21)
- {c} Disorders associated with a wide range of emotional or behavioral difficulties are

worse than a single symptom or narrow range of problems. This is why we selected the *Global Severity Index* on the SA-45 rather than using scales which measure dimensions of specific problems, such as “depressed” children, “hostile, aggressive” children.

Problems on single dimensions can be readily identified by a number of testing strategies, including the full Achenbach, the long version of the Conner’s, or the SA-45, *but even very high T-scores (over 80) on any single dimension*, such as the aggressive scale, is not a good indicator of clinical significance.

- {d} Problems out of keeping with normal developmental trends usually have a worse outcome than those problems that are severe exaggerations of age-appropriate phenomena. (This criteria is measured through the CGAS, section 1.21)

Rutter (1995) states that traditional models (such as developmental theories by Piaget and Freud) which tried to explain disorder and predict its course are no longer adequate in the light of empirical longitudinal research. The new approach, developmental psychopathology, offers a much more accurate paradigm for answering questions, such as:

- {a} which of the “distressed children” are likely to get worse or better, all things being stable
- {b} when does normal misery and misbehavior, during childhood, change into something with long term adverse consequences to the child and society
- {c} what are the possible events (therapeutic and otherwise) that will move this child on a better pathway leading to positive outcomes

The validity of the risk screening protocol is determined by its ability to answer these questions correctly (according to the state of the art on childhood disorders)

Rutter (1995) and Costello & Angold (1995) review a wide range of issues and research results from longitudinal studies. Their overall point is that there many pathways to many different disorders ranging from autism, to genetic disorders, behavioral disorders, schizophrenia and familial and social problems. Predictions about a child’s future success or trouble will be much more accurate if we incorporate a broad range of influences. For example, Rutter takes issue with the simple explanation of antisocial conduct disorder offered by theories of social learning by Patterson and others. Although, he agrees with Patterson that parental family management practices have a powerful impact for good or bad on antisocial conduct disorder, many other family processes are also powerful - and in the hands of a skilled therapist can be bring about positive change for the child. These other processes include the child’s attachment to his parents, triangulation effects, scapegoating, stressors and resiliency. Moreover, the social context outside the family (school, peers, neighborhoods) also play a powerful role shaping the future path of conduct disorders.

This implies that for risk screening to be effective predictors of the future, the instruments must cover a broad range of issues (including behavior, symptoms, fitting into society) as well as issues in the social context including attachment, family process, stress and resources. The criteria for risk screening (section 1.2) of the OARTY system covers these concepts.

Sroufe and Rutter (1984) laid out the rules for any system or body of research to predict the future with children, and by inference to quantify the risk facing any child. It is apparently easy to predict backwards - to analyze a disturbed adult and figure out how he ended up in so much trouble; however, going in the other direction is much more difficult. The presence of specific risk factors, such as a history of abuse, is highly correlated with dangerousness in very disturbed adults, but only a tiny minority of abused children end up at that extreme outcome. The risk factor, itself, is not predictive, but rather it is the *risk mechanism* or process through which the individual child transacts with his genetic endowment and social environment that produces the outcomes observed across the life span.

In a similar vein, early replicas of behavior observed in childhood (e.g. aggression) do not predict the behavioural indicators of adult pathology (e.g. violent offenders). Rather, the strongest predictors in childhood appear to be adaptational failures, defined in age-appropriate terms. (Sroufe & Rutter, 1984)

Risk screening protocols vary according to whether the target group and mandate of the service is *universal* and aimed at the general population, *selective* or aimed at population with a high risk of developing long term adverse outcomes (such as early intervention with aggressive boys) and *indicated* - where the service is directed at those already presenting serious problems and heading towards even worse outcomes. (Costello & Angold., 1995; Tolan, Guerra & Kendall, 1995a). In a companion article, Tolan, Guerra & Kendall, (1995b) review a wide body of literature on different strategies of risk assessment leading to the identification of those most in need of intervention. They emphasize a developmental-ecological framework of data to obtain empirically valid predictions and useful information for intervention.

4.0 The Multigate Screening Structure

One position is very clear from the references cited above: no single instrument or domain of inquiry will meet all of the criteria for good predictions and valuable information for intervention. On the other hand, it would not be practical, affordable or cost effective to give every case referred for service, a whole battery of tests. For this reason, many researchers have developed a powerful strategy for identifying children at risk; the new method is to test children through an organised sequence of steps or “gates”. By combining the scores on two or more tests as children go through each “gate”, the researcher is able to predict more accurately with fewer tests.

The use of *multimethod, multistage* techniques in the field of psychiatric epidemiology (Brandenburg, 1990) pioneered by Rutter, has improved the accuracy of prevalence estimates for childhood disorders. The most accurate predictions occur when investigators combined a test of the degree of behaviour problems with those which measure functional impairment (CGAS).

The concept of multiple gating has been applied extensively in prevention projects (August, 1995; Lochman, 1995; Loeber, 1990; Loeber,1987; Loeber, 1984). Multiple gating has also been reported as a screening instrument for controlling access to intensive residential services (Newman, 1987). Newman affirms many of the values and practical considerations of the OARTY system (see section A.), especially linking the risk screening and global outcome data directly in the plans of care. Newman provided a table of useful combinations of testing over a multi-gate structure. Newman recommended a combination of the Children Global Assessment Scale (CGAS), a child behaviour checklist (such as the Conners'), a measure of family functioning, and the SCL-90, which is the earlier version of the SA-45.

Most of the applications of multiple gating - cited above - have been in the field of conduct disorders and concerned with future criminality. The multiple gating approach has improved the accuracy of prediction, while minimising the number of screening devices.

However, OARTY residences serve a much wide range of children than simply conduct disorders so we applied the multiple gating rules to a much wider variety of adverse outcomes with a more diverse group of children including those who are:

- medically fragile
- autistic
- with a dual diagnosis (DH + conduct or psychiatric disorder)
- with emotional and psychiatric disorders
- with conduct disorders
- socially disadvantaged, abused and under-developed

The references cited above experimented with having two or three gates, and with different combinations of tests. The rules of multiple gating relative to risk prediction (August, 1995) are:

gate #1 *the gate 1 instruments screen for premorbid risk factors - i.e. evidence of cross-setting disruptive behaviour (in the case of antisocial conduct disorder)*

Measures of autistic symptomatology, emotional disturbance and adaptive failure were added as premorbid risk factors for other adverse outcomes.

gate #2 *the gate 2 instruments screen for a mediator variable suspected of catalysing the pathological effects*

The catalytic factors include the child's current stress levels, the parent's current

stress levels, the accumulation of serious stressors and adverse conditions on the child, the level of attachment to his or her primary care givers. (Rutter, 1995)
 As discussed below, both the premorbid risk factors (which measure aspects of health and safety) and the catalytic risk factors (which measure the child's and family's ability to cope) are predictive of escalating problems and adverse outcomes across the life span. When any of these risk factors stand alone, they are not better than 70% accurate in their prediction; however, when combined in two gates, their predictive accuracy improves dramatically - thereby lowering the chance of a false positive risk identification.

Some authors recommended using the same checklist (Conners' or Achenbach) at both gates, but obtaining the data from different sources: teachers at gate #1 and parents at gate #2. Using two sources of information on the same checklist does add some new information, but in testing this procedure, the percentage of accuracy improved marginally. (August, 1995; Lochman, 1995)

5.0 Measures of Health and Safety

The gate #1 instruments which measure the premorbid risk factors for adverse outcomes that seriously compromise the health and safety of either the child or the public are as follows:

hyperactive/inattentive behaviour	<i>Conners' Global Index</i>
emotional disturbance	<i>SA-45's Global Severity Index</i>
	<i>FAB-C Problem Index</i>
adaptive failure	<i>Children's Global Assessment Scale</i>
autistic symptomatology	<i>Children's Autistic Rating Scale</i>
nursing & physical care needs	<i>Level of Assistance</i>

Only one numbers for each test is entered into the risk screening protocol, specifically the total index T-score (Conners'), the Global Severity Index T-score (SA-45), the Problem Index (FAB-C), the Children's Global Assessment Scale (CGAS) score, the total score from the CARS and the Grand Total Score from the Level of Assistance. (LA)

5.1 Conners' Global Index

The Total Index T-score from the Conners' is highly correlated ($r = 0.82$) with the Total Problems T-score from the Achenbach. (Achenbach, 1991) Therefore, agencies can substitute the Conners' with the total problem T-score from the Achenbach if this data is available.

For certain groups of children in psychiatric hospital for serious mental health problems, the scores on the Conners' and the CGAS are inversely related. (Shaffer, 1983) Using both the

Conners' and the CGAS greatly reduces the percentage of false negatives. The Conners' has been widely used in multiple gating systems. (August et al, 1995, Ferguson & Horwood, 1995) Similar scales (SBQ) have been used in Montreal (Gagnon, 1995). The power of the *total problem scores* of various behavioural scales to correctly predict clinical serious problems is between 50% and 60% when the cut-off scores are set to above T = 72. (Lochman, 1995; Verhulst, 1995). We have set the "high risk zone" for the Conners' at T = 80 or higher, well above the threshold reported in the literature. The Conners' was chosen instead of the Achenbach for several reasons the most important being that the Conners' is much shorter (10 questions, compared to 114) and less costly. Both tests are outstanding and produce comparable information.

Several longitudinal studies have evaluated the predictive validity of both the Achenbach and the Conners'. When using the Conners' or the Achenbach as a screening device, caution must be observed since a majority of children whose T-score is above 65 and below 72 (within the *medium risk zone* of the OARTY system) will not present disturbed behaviour a year later with or without intervention. High scores on these tests merely mean that the child has behavioural traits which are highly correlated with the behaviour of a population of children with proven psychiatric illness. In the case of boys, the percentage of false positives for conduct disorder on the Achenbach was 69%. This was reduced to 59% when both the teacher and parent version were completed. (Lochman, 1995).

When the cut off rate was set to very high levels (over T = 72), the percentage of false positives dropped to 30%. Similar results were also reported in a longitudinal study conducted in New Zealand on 709 children tested at age 8 years and then at 15 years. (Fergusson/Horwood, 1995).

In a six year longitudinal study in Holland, (Verhulst et al, 1994), 946 children between the ages of 4 and 11 years, were tested every two years using the Achenbach parent and teacher forms. These children, who were drawn from the total population of a province in Holland, were followed for six years. The researchers watched for any of six bad outcomes: {a} learning problems in school {b} had been expelled from school for behaviour problems {c} had received mental health services, {d} was judged by the parents as needing professional help for which he was not receiving, {e} had shown suicidal behaviour and {f} had been in trouble with police.

The six bad outcomes studied in the Holland are far less serious than the bad outcomes which we seek to predict in the OARTY system, since almost every child placed in OARTY already one or more of the bad outcomes reported in the Verhuist study. However, the Verhuist study is still very important, because even with a relatively benign list of bad outcomes, high scores on the Achenbach predicted bad outcomes for only 36% of boys and 56% of girls. When the *attention* scale of the Achenbach was used (which is more closely related to the Conners' Global Index), a much higher percent of boys (75%) were correctly identified.

O'Donnell, Hawkins and Abbott (1995) conducted a 3.5 year longitudinal study of boys in order to study ways to improve the prediction of which boys with high scores on behaviour problems

before age 11 would get worse by the time they are teens (“worse” was defined as exhibiting drug abuse and aggression). In this study (and others referenced in the article) *more than half of boys who were aggressive before age 11 stop being aggressive by the time they are 14 years and they do not develop other behaviour problems.*

The findings of the O=Donnell study are very important for treatment of children with “high scores” on behavioural measures:

- {1} low socioeconomic status and single parent households **did not predict** who gets more delinquent with age

Paradoxically, white children from middle or middle upper income families and two parent families were more than twice as likely to become involved in drugs than children from visible minorities, lone parent families or poor families.

- {2} Once you restrict the study population to children already scoring above 70 on the Achenbach, even higher scores by teachers on the aggressive behaviour **did not predict** who gets worse or who will cease their problem behaviour two years later.

- {3} The factors which **do predict** who gets worse or better are *school bonding* (i.e. the children who like school), *skills for prosocial involvement* (including empathy, cooperation, engaging others, etc), *educational achievement*, *norms* against substance abuse and *avoidance of peers and adults involved in antisocial behaviour*. The factors that promote positive changes with children who have high scores on problem behaviour have been incorporated into the standardised plan of care model (section 1.5)

The Conners’ Global Index is very reliable. Cronbach’s Alpha measuring the internal consistency for the Conners’ Global Index was 0.89. (Conners’ Manual, 1999, page 111). Moreover, the Conners’ Global Index was highly correlated with the total score of the Kovacs Children’s Depression Inventory (CDI) at .71 (Conners’ Manual, page 131) and the total problems scale of the Achenbach $r = 0.82$ (Achenbach 1991, page 85)

5.2 Children’s Global Assessment Scale (CGAS)

The Children’s Global Assessment Scale (see section 1.2) was developed by a team of psychiatrists at Columbia University, New York. (Shaffer et al, 1983). The CGAS was developed by revising Axis 5 from the DSM-IV (Global Assessment of Functioning) so that it is suitable for children. It has since been used as a gating instrument in tertiary level services and found to be a reliable and valid indicator of the intensity of service needs. (Guzder et al, 1996; Newman et al, 1987) There is strong evidence that diagnosis and client/patient demographics are poor predictors of the costs of clinical intervention services. However, the prediction of resource

consumption and costs are increased dramatically when an intensive measure, for example, level of functioning, is considered. Newman, et al recommended the Children's Global Assessment Scale as the global tool and this to be followed by a multifaceted set of tools specific to the target population and the purpose of the program. This strategy was also proposed by Rutter (1985) Kazdin and Kagan (1994).and Brandeburg et al (1990).

The CGAS shares the psychometric properties of the DSM-IV axis 5, which has in existence since 1962 and has been subject to thousands of research studies., including extensive use by Michael Rutter, who found that this scale was predictive of life span adverse outcomes (Rutter, 1985). The inter-rater reliability of the CGAS is 0.84; test-retest correlations are in the high 80's and 90's. The CGAS accurately distinguished inpatient and outpatient populations.

5.3 The Symptom Assessment - 45 (SA-45)

This is a one page form with 45 questions completed by the teenager. The SA-45 (see section 1.21) provides a direct measure of emotional disturbance (which is measured indirectly and in a limited way by the Achenbach). The SA-45 is a revision of the BSI by Derogatis. The longer version (SCL-90R by Derogatis) is the major epidemiological instrument for adult mental health. The SA-45 produces one number for input into risk assessment and outcome measurement (the global symptom severity scale) and ten other scales that can be used in case planning. The SA-45 is employed in over 100 settings (SA-45 manual).

An index of Global Severity and a Positive Symptom total can also be obtained from the SA-45 results. These indices are helpful in assessing overall symptomatology. Appendixes in the technical manual provide inpatient and nonpatient age- and gender-based norms to which you can compare your respondents' results. The SA-45 is scored relative to a normative database of over 18,000 subjects. The technical manual also provides information on the administration of the instrument, interpretation of scores, the development of the SA-45, how to export statistical data, and other important details. The internal consistency of each of its scales has been established with Cronbach's alpha of 0.71 or greater. Test-retest reliability over a two week interval are in the 80's. (SA-45 manual, reliability summary, page 49)

The SA-45 is a self-report measure that can be administered and scored by individuals who do not necessarily have advanced formal training in clinical psychology or psychometrics, which makes this scale especially easy to administer to large groups.(SA-45 Manual, 1999)

Borduin et al (1995) found that the Global Severity Index of the SA-45 "represents the best single indicator of the respondent's psychiatric functioning" (page 572) The manual for the SA-45 provides extensive data on the reliability and validity of the SA-45.

5.4 The Feelings, Attitudes and Behaviours Scale for Children (FAB-C)

The FAB-C consists of 48 *yes/no* questions that ask the child to rate whether the statement describes him or her. The FAB-C is appropriate for children ages 6 to 13 years of age. For younger children and those with little reading ability, the FAB-C can be administered by reading the questions to the child and recording his response on the form.

The FAB-C was developed by Joseph H. Beitchman, M.D. who is a Child Psychiatrist practicing at the Clark Institute. The rationale for developing this self report measure grew out of the author's published concerns about guiding clinical practice and research solely on the basis of the observations of parents and teachers. (Beitchman and Raman, 1979) There have been extensive studies (reviewed in the FAB-C manual, pages 27-28) that have uncovered persistent informant bias with behaviour observation tests, such as the Conners'. When teachers rate behaviour, they show evidence of a "middle class bias" (Beiser, 1972; McDermott, Harrison, Schroger & Wilson, 1965), an excessive emphasis on academic excellence (Lessing, Oberlander, & Barbera, 1974) and a defective understanding of mental health (Hsu, 1966). When parents rate behaviour, researchers have evidence of both over-reporting and under reporting behaviour (Speer, 1971).

Beitchman does not dismiss behaviour observation scales, stating that they make a substantial contribution to child psychiatry and psychology. However, Beitchman points out that adding a self report measure "may lead to improved clinical methods in which the nature of a particular problem can more accurately be specified as resting as resting in the perception or behaviour of the adult, the behaviour or feelings of the child, or both. The use of self report measures also has important treatment implications, since these type of measures provide information that allows treatment to be better directed at the specific problem." (Beitchman, 1996, page 27)

Beitchman's comments about the role and importance of self report measures apply equally to the SA-45 which is a self report measure for adolescents.

5.41 Relationship of the FAB-C to other Self-report measures for children

There are other self report measures for children, specifically, the Piers-Harris Self-Concept Scale (Piers, 1969); Self-esteem Inventory (Coopersmith, 1967), the Rosenberg Self-Esteem Scale (Rosenberg, 1995), the Self-Perception Profile for Children (Harter, 1985); the Index of Peer Relations (Hudson, 1986); the Nowicki-Strickland Locus of Control Scale, (Nowicki, 1973); the Children's Manifest Anxiety Scale (Reynolds, 1980); the State-Trait Anxiety Inventory for Children, (Speilberger, 1973) and the Children's Depression Inventory (Kovacs, 1992). Each of these scales have good psychometric properties and have been in used in both clinical and research applications for decades. Many of these scales are in the public forum and can be copied without royalty fees.

The disadvantage of the aforementioned scales is that they are limited to specific areas of the

child's functioning (i.e. anxiety, peer relations, self-concept, depression, etc). These other scales do not have a global problem index and they do not cover a broad range of issues. The FAB-C on the other hand was engineered to cover: conduct problems, self-image, worry, negative peer-relations, antisocial and lie-defensiveness. In addition, the FAB-C has a total problem score.

5.42 Reliability and Validity of the FAB-C

The FAB-C was normed on a sample of 6 to 13 year old children, representative of the Canadian population (Beitchman, 1996). A total of 1,988 with a mean age of 9.6 years comprised of the sample and made up of a cross section of Ontario society by gender, socio-economic status, family structure and cultural background. Separate norms for males and females and four age-groups (6-7 year olds, 8-9 year olds, 10-11 year olds and 12 to 13 year olds).

Cronbach's Alpha, as measured by the Kuder-Richardson formula 20, varies from .55 to .79. These reliability measures are concerned with the internal consistency of the test items in each scale. Internal consistency reliability measures errors that arise when the specific items on the test are not representative of the larger conceptual domain that the test is trying to measure. (Kaplan, 1997, page 125) A reliability score in the range of .70 to .80 is required for most research. The most reliable scales of the FAB-C are those which measures conduct problems, lying and negative peer relations. Negative peer relations is one of the two most important predictors of life span outcomes (Rutter and Sandberg, 1985).

Overall the FAB-C has adequate internal consistency. The FAB-C manual provides additional detail on reliability including test-retest reliability. The manual also contains extensive data on the validity of the scales.

The FAB-C displays low correlations (<.30) for most scales of the Achenbach and Connors' behaviour checklists, which is similar to the pattern with the CGAS. This is good because these tests measure similar but distinct constructs (as discussed in the rationale section 5.4). Therefore each of these tests adds new information; they are not just substitutes for one another.

A large group of children (403) were selected across Ontario and tested in a variety of ways to ensure that they were normal. Another group of children (103) were selected from the inpatient and outpatient services of a large Psychiatric Hospital. The entire group were tested on the FAB-C and the problem index correctly classified 71.4% of the children. The degree of discriminant validity for the FAB-C is similar for other tests in threshold {a} and is considered quite acceptable. The fact that about 30% of children are incorrectly classified by any single test illustrates why we need multiple measures and two thresholds to accurately assess the risk.

The relation between the FAB-C and various Psychiatric Diagnostic measures and the Childhood Depression Inventory (CDI) were clearly demonstrated. For example, 78% of children with clinically elevated scores on the CDI were identified by children with a T-score of 65 or greater in the problem index (excluding children with elevated scores on the lie scale).

5.5 Children's Autistic Rating Scale

The Childhood Autism Rating Scale (CARS) is published by

Western Psychological Services
12031 Wilshire Blvd
Los Angeles, CA, 90025-1251

The authors Eric Schopler, Robert Reichler and Barbara Renner developed the instrument in 1971; since that date, the CARS has been used in hundreds of studies. The internal consistency reliability (measured by coefficient alpha) is .94. "This measure indicates the degree to which the scale, taken as a whole, measures a unitary phenomenon rather than numerous unrelated facets of behaviour, and it provides the justification for combination of the 15 individual scale items into a single total score. It is from this total score that a classification is determined." (Schopler, 1992)

Extensive data is provided on all aspects of reliability and validity as stipulated by the American Psychological Association and other associations concerned with test standards.

The CARS has also been validated "when used as a screening tool by a variety of well informed individuals who were not necessarily psychodiagnosticians" Various studies have shown the reliability and validity of this application with various professionals who had only minimal training and exposure to autism and had been trained on the administration of the CARS by videotape instructions. (Schopler, 1992, manual page 6)

The validity of the CARS has been demonstrated in numerous research studies and is listed as a mandatory screening tool for the Intensive Early Intervention Program for Children with Autism.

5.6 The Level of Assistance Scale (LA)

There are three parts to the Level of Assistance Scale:

(1) *special needs of a medical nature*

This is a series of medical conditions that require daily intervention or management to enable the child to function in various aspects of his physical self: neurological, musculoskeletal, skin, gastrointestinal, sensory, respiratory and psychosocial. This structure is similar to the Sickness Impact Profile or SIP, which is used in many hospitals. (Kaplan, 1997, page 522)

Under each of the above areas, there are a series of conditions. Answer "yes" if the condition is present as in the example below:

Step one: Identify any special needs of a medical nature

AREA OF SPECIAL NEEDS	SPECIFIC CONDITION	YES OR NO
<i>Neurological</i>	Seizures	
	Cerebral Palsy	
	Acquired Brain Injury	
	Microencephaly	
	Hydroencephaly	

The raw score for part one is simply the total number of "yes" answers.

(2) *the degree of staff support required for the child to function physically*

The degree of support is a scale that varies from totally independent (zero) to total care (4) as illustrated below:

:

LEVEL OF CARE	SPECIFICATION	SCORE
independent	no special needs compared to children of their age	0
daily oversight	Child requires daily supervision	1
verbal prompt	Child requires verbal prompting	2
physical prompt	Child requires hand over hand guidance	3
total care	Staff must do everything	4

There are five physical areas (positioning the body, mobility, manipulation, oral motor and communicating needs) that are measured in relation to this scale. This part of the LA is based on the Katz Index of Activities of Daily Living. (Kaplan, 1997, page 524) A score from 0 to 4 is assigned to each area. The raw score for part two is the sum total of all scores.

(3) *The degree of support required for the child to function socially*

The degree of support scale is then applied to 13 areas of social and communicative functioning. The areas and specific skills in this part are adapted from the Syracuse Community Referenced Curriculum for Students with Moderate to Severe Disabilities. (The Syracuse ... 1989)

This area is scored in the same manner as part two above, by assigning a score from 0 to 4 for each skill. The raw score for part three is the sum total of all scores.

Measuring Risk with the Level of Assistance Scale

The grand total score is entered in the risk screening system. The grand total score is a composite of the three parts, calculated in the following manner:

Step Four: Compute the composite score

AREA MEASURED	INSTRUCTIONS	SCORE
Medical Special Needs	Multiply total number answered <i>yes</i> times 4	
Support for Physical Functioning	Insert total score	
Support for Social Functioning	Insert total score	
Grand Total Score (sum of all scores)		

Low Risk = 12 or less
 Medium Risk = 13 to 30
 High Risk = 31 or more

- The concept of *low risk* means that the child has special needs but can live a normal life with adequate support (such as medication, assistive devices, special education and family support).
- The concept of *medium risk* means that the child’s special needs in areas of physical functioning, nursing care and social functioning are so great that he or she will probably never live outside a protective setting or to live without the ongoing supervision by a care provider.
- The concept of *high risk* means that the child cannot function or even survive without continual assistance, supervision and intervention.

The Level of Assistance scale is administered by the child's caregiver or a worker with detailed knowledge of the child's special needs. Infants and children who are medically fragile are the primary candidates for this test; however, other children with chronic medical conditions should be tested on this instrument. Higher scores on the LA should reflect a higher workload on the primary caregiver and this instrument may be used to assess the burden of care.

Validity of the LA

The Level of Assistance Scale (LA) is a completely different type of test compared to others in threshold {a}. The LA is a *criterion referenced test*, as opposed to a norm referenced test. A criterion referenced test compares performance with some clearly defined criterion for mastery. Criterion-referenced tests are very common in medical care to measure quality of life. The test used by OARTY is made up from parts of several existing medical and developmental tests.

The validity of the LA is acquired indirectly from the fact that it is a list of diagnostic categories and medical conditions, based on medical reports from Ontario physicians and hospitals. The validity of the LA is derived from the validity of the diagnostic procedures that provide the data.

5.7 Conclusion on Premorbid Risk Factors

In summary, the six instruments employed at gate #1 (Conners' Global Index, the CGAS, the SA-45, FAB-C, the CARS and the LA) are proven instruments employed in thousands of research studies around the world. The manuals for the SA-45, the CARS and the Conners' review the reliability and validity of these instruments, including their predictive accuracy. Shaffer (1983) reviewed the reliability and validity of the CGAS. The literature on developmental psychopathology (section 4.0) has found that the concepts measured by these four instruments are the best premorbid risk factors for very bad life span outcomes. All instruments at threshold {a} measure the premorbid risk factors at the interval level and therefore the full set of parametric statistics can be used to study the patterns.

6.0 Measures of the Child's & Parent's Ability to Cope

The central belief behind the second threshold is that the *prognosis for a child* with serious problems in the areas of mental health, fitting into society or the child's behaviour is determined by a combination of

1. The child's *current stress levels*
2. The parent's *current stress levels*
3. The *accumulation* of very stressful events and adverse conditions from early life
4. The quality of the child's *attachment* to his or her primary care givers
5. The child *acquisition of basic skills* in communication, socialisation, daily living and motor development compared his peers

These concepts are measured by:

the child objective stressors checklist the child's experience of stress
the QRS-F..... the parent's experience of stress
a sociodemographic checklist..... adverse family history
Parental Bonding Instrument..... quality of attachment
Vineland Adaptive Behaviour Scale (VABS) child's developmental functioning

6.1 the Objective Stressors Checklist

This is an open ended one page questionnaire to help a child identify the stress in his life from his point of view. It clearly draws the child into his plan of care. As the child reveals information about "what he is worried about", the worker will have to engage the child in a discussion about "how we can help". The results of this dialogue will be incorporated into the plan of care.

The actual number which feeds into the risk screening system is the “number of discrete stressors identified by the child”. This validity of this strategy for measuring stress has been discussed in the literature. Quoting from Blair Wheaton:

“Basically, the items on stress indices are not supposed to function as correlated measures of a single underlying concept of overall exposure to stress. Rather, the assumption is that each additional stressor stands for a separate increment in the level of environmental demand or complexity the individual must face.” (Wheaton, 1983, page 214) In other words, the stressors checklist - and all other indices of stress - should be not be treated as if they were composite tests such as the Conners’ Global Index”, whose validity is derived from the fact that the 10 questions contained on the test are highly correlated with a common “target concept”, which in turn is predictive a long term behaviour problems.

Moreover, there is no dimensionality in the objective stressors checklist. The child either identifies or does not identify one or more stressors. This also eliminates the problem of reliability. The test examines the question, “does the child perceive a stress in such and such area: yes or no, and if yes describe the stress in the child’s words”. There is no common ground for any inter-rater involvement and reliability over time has no significance either, because any changes over time reflects real changes in the child’s perceptions, not bias in the first measurement.

There are only two threats to the validity (and reliability) of the instrument. The child may not identify all of the things he is worried about - possibly because he is too worried about expressing them. It is also possible that he may overtly lie by identifying something as a stress that he is not worried about. The first threat is quite plausible with younger and immature children; if they cannot talk about their worries, this will produce a false negative result. Workers will have to be aware of this danger and employ the clinical over-ride option on the risk classification system.

The second threat (overt lying about things that worry them) seems unlikely since there isn’t any reinforcement or benefit to the child in doing so. However, if the worker has reason to believe that the child is exaggerating his worries, then he or she will use the clinical over-ride option.

Rutter noted that there are two very distinct groups of children with disorder which are visible in standard epidemiological data: (a) childhood problems which are situation specific and (b) childhood problems that occur in several different situations, persist over time and are associated with a wide variety of emotional and behavioural difficulties. Situation specific problems, even if they are quite severe and disabling at one point in time, respond quickly to changes in their interpersonal relationships. Many childhood disorders are situation specific. The worst prognosis and the greatest need for intervention is indicated by group (b), the children with broadly based difficulties. These children are most often associated with severe environmental stressors. (Rutter, 1985)

The sheer number of objective stressors that people self-identify has been shown to be a strong predictor of physical abuse by parents (Chan, 1994), Kasim, 1995; Kruttschnitt, 1994). Stress has long been identified as a factor in the generation of physical illness (Wheaton, 1983) and mental illness, especially depression, substance abuse and antisocial conduct disorders. (Dohrenwend, 1992)

Researchers have found that *life events stress* and *neighbourhood violence stress* along with normative beliefs supporting violence *predicted violence* much better than poverty in a two year longitudinal study involving 1,935 multi-ethnic elementary school age children from lower income inner city neighbourhoods. (Guerra, 1995)

In a study of the separate effects of different types of stress, Blair Wheaton compared life events, chronic stress, non-events, daily hassles and childhood traumas. Because of the nature of the study, it was impossible to include macro stressors. However, *the study concluded that each separate category of stress made an independent contribution to mental health problems*. He found that {1} no single stress concept was sufficient to capture the full predictive power of stress; each category of stress makes a separate and independent contribution to the bad outcome {2} *life events stress* (family moving, parents splitting up, etc) do not make a smaller contribution to bad outcomes; {3} the explained variance attained by adding the other stress concepts was three to five times greater than using life events stress alone and at times it approaches 50%, *meaning that the full continuum of stress explains half of the factors leading to mental illness*; {4} non-events and daily hassles played a generally weak role in predicting mental health outcomes, but their effects were not simply absorbed by the other stressors - they made a separate contribution; and {5} estimation of the true role of stressors clearly demands a model of the effects of cumulative stress exposure over time. Sources of stressors which persist over a longer period of time are more toxic, partly because of the chain of events flowing from a serious stressor. In studies of adult mental health, childhood trauma had the second greatest total impact on adult mental health after current chronic stressors. (Wheaton, 1996)

The concepts identified in the objective stressors checklist were taken from the above article by Blair Wheaton. Technically, some of these stressors (such as serious traumas and chronic stressors) should be given more weight, but since the value of the weighting cannot be estimated at this stage, we will give all articulated unique stressors the value of “1”. This protects against a false positive or inflated value of the objective stress load on the child.

6.2 the Questionnaire on Resources and Stressors (QRS-F)

The QRS-F is 52 item checklist of family stressors specifically validated on families who were caring for handicapped children. The QRS-F produces a total stress score and four sub scores as:

- Parent and family problems
- Parental pessimism

Parent perception of the child's problems
Parental perception of the child's physical incapacitation

Research on families with handicapped children indicates that these families experience much higher levels of stress than other families in the community. (Perry, 1992)

Research into different groups of abusing and non abusing parents found that: (a) stress is the most important correlate of child abuse potential and other secondary factors which themselves contribute to child abuse including social support and family resources. Both the QRS and the PSI were found to accurately assess the child abusing levels and types of stress (Burrell, 1994).

Holroyd (1983) reviewed 48 studies using the QRS, which give evidence of its reliability and validity. Friedrich, Greenberg & Crnic who created the version of the QRS we are using at OARTY reviewed the psychometric properties in their article (Friedrich, 1983) The Kuder-Richardson 20 or KR20 reliability coefficient for the QRS-F is .93. The QRS-F has a mean value of 18.6 and SD of 11.0).

The criterion validity of the QRS-F was established as follows:

the parent/family problems scale	Beck Depression Inventory (r = .41) Marlowe Crowne Social Desirability Scale (r = -.38)
the pessimism scale	Beck Depression Inventory (r = .39)
the child characteristics scale	problems scale (r = .43)
the physical incapacity scale	problems scale (r = .67)

These results and the research cited are all excellent foundations for the use of the QRS-F.

6.3 the Socio demographic checklist

This is a 15 item checklist of serious risk factors in the social background of the children. The actual number of risk factors is entered into the risk screening system.. A review of the literature for this checklist revealed that having four or more of any single item in the background meant that the child had better than a 66% probability of being unable to function independently by age 18 years. (Werner, 1989 and 1992)

Kazdin and Kagan (1994) reviewed the literature on risk factor research. They concluded that:

- {a} No single risk factor may be essential (in producing a disorder), but rather a combination of factors is pathognomic, for example, marital discord, criminal behaviour of father, harsh child-rearing practices, early signs of aggressive behaviour, and lack of parent supervision predict the onset of conduct disorder.

- {b} The presence of only one of these (as itemised above) risk factors does not increase the risk for later conduct disorder, if the others are not present.
- {c} When two of these risk factors appear together, the probability of conduct disorder increases fourfold.
- {d} When three or four factors are present, the risk increases several times more.

In applying these findings, we selected the most conservative thresholds designed to minimise the false positive potential. Accordingly, we determined that zero or one risk factor was in *the low risk zone*, two or three risk factors were in the *medium risk zone* and four or more risk factors were in the *high risk zone*.

The sociodemographic checklist is a very small list of the background factors which are robust risk indicators, whose effects endure across all developmental stages, gender and ethnic identity. Many risk factors from epidemiology studies have not been included on the basis of recent publications from such eminent authorities as Michael Rutter, Jerome Kagan and Alan Kazdin who have raised serious questions about the validity of these assumptions of risk, especially if these factors are to be used as criteria for screening clinically significant cases for access to service. Many of these factors from the child's family background which were correlated with disorder in epidemiology studies have been found through longitudinal research to be active only when they occur in extreme degrees or only in certain developmental stages. (Kazdin, 1995; Rutter, 1985) We have excluded these unstable risk markers to reduce the danger of false positives.

The sociodemographic checklist was compiled from those risk factors that met the following criteria:

- {a} The presence of the stressor or condition is universally toxic to all children regardless of ethnic background, age or gender
- {b} The destructiveness of these factors remain active throughout the life span of childhood on the most common disorders.
- {c} There is a fairly high prevalence for each of these risk factors (between 1% and 5%), so that referrals from CAS agencies would likely display at least one of them.

This list does not include risk factors that are active only during the early childhood years (eg difficult temperament, birth weight under 2500 grams, low Apgar score, duration of breastfeeding, maternal smoking during pregnancy, and childhood lead intake). These risk factors are of great importance in primary prevention, but have little discriminating power in assessing the clinical significance of a referral for clinical services. (Dishion, 1995; Fergusson, 1995;

Kazdin, 1995; O'Donnell, 1995; Rutter, 1985; Andrews, 1992).

Other examples of unstable risk markers include age of onset, (Tolan, 1995), single parent, family income level and urban residence (Dishion, 1995; Guerra, 1995; Vaden-Kiernan, 1995) and finally, family dysfunction and reading problems (Rutter, 1980). However, we did include as a risk marker, physical and sexual abuse and learning problems starting at elementary school - which are persistent and reliable risk markers. (Tremblay, 1995)

Sameroff (1995) reviewing 14 longitudinal studies as well as his own research found that the continuity of risk factors, such as:

- abusive and unemotional parent-child interactions
- family conflict and distorted levels of communication
- lack of social support
- stressful life events
- poverty
- disorganized neighbourhoods
- inadequate school systems

in the context of a child is much stronger than the continuity of any behavioural domain within the child. (Page 688) Whereas the correlations for adaptive behaviour scores at 4 and 18 years of age was .40, the correlation for environmental risk across the same 14 - year period was .80. Given the transactional nature of the risk mechanism between the child and his context, the overwhelming continuity of risk factors is what drives the correlations between early adaptive failure and adult psychopathology.

For a complete review of the items included and excluded from the socio demographic checklist, please see the accompanying paper, Sociodemographic Risk Factors.

The number of items checked on the sociodemographic checklist is correlated with the number of items articulated on the objective stressors checklist at $r = .53$. Both the continuum of stressors and the sociodemographic checklist are highly correlated with the SA-45 ($r = .72$, Sociocheck) and ($r = .65$ continuum of stressors).

6.31 How is the sociodemographic checklist of risk factors different from the objective stressors checklist?

The source of data for the socio demographic checklist is the referral information, which reviews the history of the child. Although, the items on the list are intrinsically stressful to the child, this list is different than the objective stressors in three ways:

- The sociodemographic checklist is a list of stressors that refer to the past and the family background of the child (such as a family member committed suicide) which the child

may not even think about currently, but which nevertheless increases the long term risk of suicide for the child by eight fold. (Davidson, 1991)

The sources of the sociodemographic checklist are the file and referring worker. The source for objective stressors checklist is strictly the child.

The sociodemographic checklist is a list of risk factors that the worker checks off as present or absent; whereas the objective stressors checklist is an open-ended questionnaire, with prompts on the type of stress only.

6.4 the Parental Bonding Instrument (PBI)

The PBI has been used extensively in cross sectional and longitudinal studies across the world. Norms now exist for converting the raw scores to standard measures. The PBI measures how securely the child is bonded to whatever target person is identified on the test. The target can include foster parent, child care worker or family of origin. The security of attachment is a critical protector of the life span outcome of many social & demographic risk factors (Section 6.3) and the prognosis for change in children with symptoms of emotional disturbance or behaviour problems. (Mallincroft, 1992; Werner, 1989 & 1992, Fonagy, 1993, Rutter, 1982; Rutter, 1995; Costello & Angold, 1995)

The Parental Bonding Instrument (PBI) was developed by Gordon Parker, Hilary Tuping and L.B. Brown (Parker et al, 1979). The original purpose of the PBI was to examine the parental contribution to a parent-child bond and to define and measure the constructs of significance.

The original validation sample was drawn made up of 150 mothers and 148 father who were medical students, psychiatric nurses, technical college students and parents from a local school. A second sample of 500 parents were selected from the patients of three general medical practitioners in Sydney. The three general practitioners were chosen to ensure a reasonable representation of all social classes.

Test-retest reliability was computed. A Pearson correlation coefficient of 0.761 was obtained for the care scale and a Pearson correlation coefficient of 0.628 was obtained for the over protection *scale*. The split-half reliability (measuring internal consistency) was 0.879 for the care scale and 0.739 for the over protection scale. The inter-rater reliability for the care scale was 0.851 and .688 on the over protection scale.

The inter-rater reliability was measured by using two raters (the authors) who interviewed the parents (independently) and scored the form based on these interviews. The inter-rater reliability of the care dimension was 0.851 and for the over protection scale 0.688.

The scales were originally developed to measure the parent's ability to bond to their children based on how well they were bonded to their own parents. There are four types of parental

bonding:

- high care-low over protection (optimal bonding)
- low care-low overprotection (absent or weak bonding)
- high care-high overprotection (affectionate constraint)
- low care-high overprotection (affectionless control)

The PBI has been used with adolescents as a means for them to rate the quality of bonding with their parents. (Rey, 1989) This study found that the PBI may be “a measure of bad parenting which is perceived by children as affectionless and controlling”. The study found significant differences between children with a DSM-III-R diagnosis of conduct disorder (n=63) and oppositional disorder (n=49) as compared with a sample of normal adolescents (n=62).

Rey reviewed the findings of the PBI in other studies in the USA and Australia involving thousands of adolescents. By 1997, the PBI had become a widely used measure of parenting with adolescents (Murphy, 1997). Murphy used the PBI with a sample of 583 US and 236 UK students. Murphy reviewed research using the PBI in Australia, UK, USA and Spain. Canetti (1997) employed the PBI with 847 Israeli students. In this study the PBI was completed along with the Brief Symptom Inventory (BSI), General Well Being Scale (GWB), the Perceived Social Support Scale (PSS) and the Social Desirability Scale (SDS). The BSI is the predecessor to the SA-45. Reviewing evidence from adult studies Canetti states that the data suggests “a strong link between the dimensions of the PBI and (*adult*) psychiatric conditions” (Canetti, 1997, page 383). Canetti=s research focussed on the relationship between dimensions of the PBI and measures of mental health in normally functioning adolescents. The students were selected from a cross section of youth in Israeli society and provides a sound normative sample.

In Canetti=s study the internal reliability (using Cronbach=s alpha coefficient) was computed. All four scales were found to be reliable with the following alpha values: maternal care = 0.75, maternal control = 0.82, paternal care = 0.80 and paternal control = 0.83.

The norms employed in the OARTY were derived from this sample, which proved to be quite consistent with previous world wide research involving thousands of children. Canatti=s conclusion is as follows:

“The strong connection between parental bonding and psychopathology found in the present study provides support for Bowlby's attachment theory. Adolescents who reported high care and low control (optimal bonding) reported less psychiatric symptoms and distress. They had a positive feeling of well being and felt supported by family and friends. On the opposite end, the combination of low care and high control (affectionless control) gave rise to psychological symptomatology and a lesser-feeling of well being. These individuals also experience less support from their surroundings”.

One of the most important studies using the PBI was published in 1992, conducted by Brent

Mallincroft (1992) at the University of Oregon. Counsellors often assume that a client's social support deficits are due to insufficient resources in their environment. This study examined the hypothesis that poor parental bonds in childhood lead to adult social competency deficits that are in part responsible for low social support. The PBI was used to measure parental bonds with 253 undergraduate students. Results indicated that poor parental bonds were related to low levels of self efficacy especially social self efficacy. This is very important for the OARTY research since it predicts social skills ($r = 0.54$) which in turn as demonstrated through multiple regression statistics leads to poor support. Indeed this pathway from poor parental bonds to poor social support accounts for minimally 30% of the variance in perceived social support. Furthermore, as evident in Cameron's recent book on child maltreatment, low levels of social support predict many bad life span outcomes, such as substance abuse and irresponsibility (Cameron, 1997).

6.5 the Vineland Adaptive Behaviour Scale (VABS)

The Vineland was originally published in 1935 and today is one of the most widely utilised measures of adaptive functioning. There are three versions of the VABS. The most commonly used is the *classroom edition*. A total of 6,000 students across the USA from all age groups provided normative data and coefficients of reliability and validity. In addition, supplementary norms were derived for the following sub groups:

- 2,200 *mentally retarded* adults of various types in residential care
- 250 emotional disturbed children in residential care
- 200 visually handicapped children in residential care
- 300 hearing impaired children in residential care

The Vineland is applicable to **both handicapped and non handicapped individuals from birth to adulthood**. The median split half reliability of the adaptive behaviour composite (the score that feeds into the risk structure) was 0.94 (using the Spearman-Brown formula). This coefficient measures the internal consistency of the scores. For the sub group of disturbed children in residential care the reliability was 0.99. Test-retest reliability and inter-rater reliability for all scales is in the range of 0.98 to 0.99. The manual provides pages of detail on the validity of the instrument. (Sparrow, 1984, the Vineland Manual)

6.6 Conclusion on Measures of the Child's or Family's Ability to Cope

The instruments above (6.1 to 6.5) constitute the second threshold of risk assessment. In effect, the issues measured help to predict whether problems identified at *threshold {a}* will get worse or better over time without intervention. The instruments at *threshold {b}* screen for a mediator variable suspected of catalysing the pathological effects of premorbid risk factors, specifically, extreme hyperactive-emotional labile behaviour, social role functioning, emotional disturbance and autism. The numbers flowing from the instruments in threshold {b} are at the interval level.

As a result the full set of statistical procedures can be brought to bear on the task of analysis.

Improvements in the child's adaptive behaviour (on measures of self care, daily living, communication, socialization and motor skills) are good *leading* indicators that the level of severity of medically fragile and autistic children is improving; although we make no assumption that improvements in adaptive functioning level actually *causes* the change in autism or medical fragility.

The hypothesis that reduction in parental distress will help to reduce levels of severity in autism is somewhat controversial, given the long history of "mother blaming" that occurred in early theories of autism. The pathological conditions of medically fragile children clearly will not change as parental stress improves. However, a reduction of parental stress levels is still a worthwhile outcome for both autistic and medically fragile children.

6.7 The Full Set of Screening Instruments

The tests listed above in thresholds {a} and {b} have been in existence since the 70's and 80's. They have been used in 100's of published research studies ever since, producing extensive evidence of their reliability and validity as predictors of bad life span outcomes.

- {a} Children's Global Assessment Scale (CGAS)child's social impairment
- Conners' Global Index child's behaviour problems
- SA-45 teenager's psychiatric symptomatology
- or FAB-C school-age child's emotional symptomatology
- The Childhood Autism Rating Scale (CARS)..... degree of autistic symptomatology
- Level of Assistancedegree of care required for medically fragile children

- {b} the child objective stressors checklist the child's experience of stress
- the QRS-F..... the parent's experience of stress
- a sociodemographic checklist.....adverse family history
- parental bonding instrument..... quality of attachment
- Vineland Adaptive Behaviour Scale (VABS) child's developmental functioning

With the exception of the sociodemographic checklist (which is a list of historical events in the child's life and therefore does not change) each instrument has been used to measure of outcomes. The reliability coefficients (internal consistency, test-retest and inter rater reliability) of each of the test selected are all above 0.70 and many are above 0.95. Five of the tests (Conners' , SA-45, FAB-C, CARS and VABS) are published by test corporations and come with extensive manuals and research on the reliability and validity of the instruments.

6.71 Universal instruments

All children will be measured by the CGAS and the sociodemographic checklist (in other words, these are universal OARTY tests).

Different client groups will use different combinations of tests in addition to the universal tests.:

6.71 medically fragile

the QRS-F..... the parent’s experience of stress
Vineland Adaptive Behaviour Scale (VABS)..... child’s developmental functioning
Level of Assistancedegree of care required for medically fragile children

6.72 autistic

The Childhood Autism Rating Scale (CARS)..... degree of autistic symptomatology
the QRS-F..... the parent’s experience of stress

6.73 dual diagnosis (DH + conduct or psychiatric disorder)

Conners’ Global Index child’s behaviour problems
the QRS-F..... the parent’s experience of stress
Vineland Adaptive Behaviour Scale (VABS)..... child’s developmental functioning

6.74 conduct and emotional disorders

Conners’ Global Index child’s behaviour problems
SA-45 **** (or FAB-C)..... child’s psychiatric symptomatology
the child objective stressors checklist the child’s experience of stress
parental bonding instrument **** quality of attachment

6.75 socially disadvantaged, abused and under-developed

Conners’ Global Index child’s behaviour problems
the child objective stressors checklist the child’s experience of stress
parental bonding instrument **** quality of attachment
Vineland Adaptive Behaviour Scale (VABS)..... child’s developmental functioning

**** for adolescents only

The risk screening system can run quite successfully using only the two universal instruments, although most children will be given a total of four or five tests.

7.0 Outcome Measurement

By retesting the child on the screening instruments *at discharge* and *every nine months in care* (if the child is in care longer than nine months), the agency and OARTY obtains valid and reliable sociometric data on changes over time in the degree of premorbid risk factors and the child's and family's ability to cope.

There are three additional levels of data collection for measuring outcomes:

- On admission and every nine months thereafter (and at discharge), the worker, the child and his parents together complete a checklist entitled *Indicators of Success*. The checklist is made up of {a} seven vision statements, {b} life span outcomes related to each vision and {c} indicators of success. The indicators are the standard service results of OARTY agencies. Children and parents assess the status of all possible indicators in relation to four categories:

S or strength - the child functions very effectively

OK the child has no problems or frustrations functioning

NW or needs work - the child has not had a chance to learn the skills or is frequently frustrated while attempting it; but this area is reserved for future work

PR should be a priority for the upcoming plan of care

- We will collect supplementary test data (both before and after intervention) for selected target groups. The target groups might be conceptualized as children with common goals or children with a specific diagnosis (such as autism) or a specific service need (such as preparation for independence).
- A plan of care model which collects information about barriers, opportunities and resource inputs (who does what to whom and with how much) in relation to each service result marked as a priority (PR) has been implemented by means of a computerized case management system installed in all agencies conducting the research

7.1 Indicators of Success

The indicators of success is a catalogue of the skills, attitudes and resources necessary for the child to adapt to his or her environment. This is the definition of a functional assessment (Brown, 1984) When a worker assesses a young person on this catalogue of skills, he is conducting a functional assessment.

A functional assessment predicts success in the community much better than symptoms, diagnosis, history of behaviour problems, traits or intelligence. (Dellario, 1984)

The structure of this catalogue (vision statements, life span outcomes and skills/accomplishments) was based on the work of Peter Pecora with special treatment foster care in San Francisco (Traglia, 1997).

The *Indicators of Success* catalogue is an example of a criterion referenced test. Criterion-referenced tests have items designed to match certain specific instructional objectives. For example, if the objective of some educational program is for children to be able to list 75% of the countries of Europe, the criterion-referenced test could ask that the countries be listed. Children who listed 75% of the countries would pass the test. They would be evaluated against this particular criterion rather than on the basis of how they perform relative other students. Validity studies for the criterion-referenced tests would compare scores on the test to scores on other measures that are believed to be related to the test. (Kaplan, 1997). The validity of each section of the catalogue of goals is reviewed in section 8.0.

7.2 second level - individualised outcome measurement

Other assessment tools will be completed by consulting psychologists, hospitals, community agencies and school boards. This individualised test data will be collected on the OARTY computerised plan of care database. By correlating this additional test data with the child's skills and accomplishments (as measured by the *Indicators of Success* catalogue), we will be able to validate the catalogue both for the individual child and at the group level.

In addition, sub groups of children (selected because they share a DSM-IV diagnosis, or a common priority area on the plan of care), will be given supplementary tests which provide an independent measure of the degree to which the child is making progress on his goals. The supplementary tests are as follows:

7.21 the Social Skills Checklist (Gresham and Elliot)

The Social Skills Checklist is a measure of social skills (empathy, cooperation, assertiveness, self control and responsibility). There are separate booklets for children from kindergarten to high school and each booklet has about 39 questions. The scoring sheet is attached and it can quickly

scored and interpreted by a foster parent or staff. This test is positive in content and the results can be readily used in casework with the child. It has been used extensively in outcome research around the world (Gresham & Elliott, 1990). The authors are among the most prominent world class experts in social skills.

7.22 the WRAT-R and the K-TEA

These tests measure the child academic functioning (reading, writing and math), producing grade equivalents. They have been used extensively in research around the world and have excellent psychometric properties (Kaplan, 1997).

7.23 the Interpersonal Support Evaluation List (ISEL)

One of the vision statements on the *Indicators of Success* catalogue is

The child will have the support and acceptance of his peers. Community centres and neighbours will be positive towards the child especially by including him in their daily lives.

The concepts measured by the ISEL are:

- tangible or material resources
- people to do things with
- people who can give advice
- people that say “you are OK” and provide emotional support

The ISEL was initially developed in 1983 by S.Cohen and H. Hoberman. The Alpha coefficients measuring the internal reliability ranges from 0.88 to 0.90. The test-retest reliability was 0.87. These results make the ISEL a very reliable measure of social support. The ISEL was validated through seven studies. “Increases in the ISEL score are consistently associated with decreases in symptomatology”. (Cohen, 1984, page 81) “Moreover, it (ISEL) has been consistently found to interact with stress measures in a manner that is consistent with the hypothesis that social support protects people from the pathogenic effects of stressful events”. (Cohen, 1984, page 89)

7.24 The Coping Inventory for Stressful Situations (CISS)

The CISS, a 48 item questionnaire, is a measure of types of coping and has three basic scales, task oriented, emotion oriented and avoidance oriented coping; the avoidance component has two sub scales, distraction and social diversion. Its psychometric properties are quite good (Kaplan, 1997). The CISS is published by MultiHealth and the manual has details on reliability and validity.

7.3 the Plan of Care Model

Every agency is required, as a condition of its licence, to prepare a plan of care within 30 days of admission, to monitor the plan monthly and to review it completely every three months. Moreover, the child, his parents, CAS worker and community partners (such as teachers) *must be invited* to participate in the initial plan and in the quarterly reviews. The plan must be recorded and copies sent to all parties.

The full plan of care includes details on family visiting, the child's education, child welfare court, medical and dental care, important events in the child's life, the child's behavior problems and the consequences and actions required to meet any other need of the child or his family.

The plan of care is a legal requirement of the licence to operate residential care. Goals of service must be set within 30 days of admission and reviewed every 3 months thereafter. All parties, including the child, parents, CAS workers, foster parents, etc must be invited to participate and must receive copies of the plan and related documents. The plan of care is like a contract of service. The assessment of goals and the process of selecting priority areas is carefully reviewed and discussed by all parties. When the child's goals are assessed as completed in this context we can assert that the assessment is at least shared by a wide group of stakeholders most of whom are not working for the residential care provider.

The plan of care includes an individualized treatment planning component, which is made up of the following parts:

- E.a Assessment of the child's needs and reasons for service
- E.b Priority Service Results of the Plan of Care
 - including details on*
 - E.b1 barriers
 - E.b2 strengths
 - E.b3 goals
 - E.b4 service diary

The goals are the stages that a child goes through in order to accomplish a milestone or learn a skill (the standard service results of OARTY). The goals are conceptualized by breaking down the learning task into small, fully achievable steps for the child to take. These goals should be defined in clear behavioral terms; it should be easy for the child and everyone to know if the child was succeeding in terms of his goals. The goals will be unique for each child, since every child will have a different pathway to their learning objective, depending on their stage of development, barriers and strengths.

The plan of care model is an application of behavioural assessment methodology. Kaplan (1997) reviewed the validity of this structure for conducting research:

The production and review of the plan of care involves the observation of behavior. In behavioral observation studies, the observer plays a more active role in recording the data and therefore is much more likely to make errors. Some of the problems include: reactivity, drift and expectancies.

7.31 Reactivity

The reliability of observers in behavioral observation studies is usually assessed in selected sessions, during which an experimenter observes the observers. Studies have shown that reliability and accuracy are highest when someone is checking on the observers. This increase in reliability is called *reactivity* because it is a reaction to be checked. In one study, observers rated behavior recorded on a videotape under one of two conditions. First, the observers were told that their ratings would be checked against a standard for accuracy. Later, the observers were told that there was no standard. In both cases, there actually was a standard against which the accuracy of each was checked. The data demonstrated that accuracy dropped by 25 percent when the observers were led to believe that their observations would not be evaluated. Indeed many studies have demonstrated that accuracy and inter-rater agreement decreased when observers believe their work is not been checked. (Kaplan, 1997, Page 204)

In general, you should always use caution in interpreting reports on inter-rater reliability. Often the estimate of inter-rater reliability is based on assessment during training. When observers are not observed [when they are actually collecting data], their accuracy will likely drop.

The accuracy (i.e. reliability) of the plan of care model will vary according to how well the process is itself reviewed.

7.32 Drift

When trained in behavioral observation methods, observers receive extensive feedback and coaching. After they leave the training sessions, though, observers have a tendency to drift away from the strict rules they follow in training and to adopt idiosyncratic definitions of behavior. One of the most common problems, *the contrast effect*, is the tendency to rate the same behavior differently when observations are repeated in the same context. Further, standards may shift, resulting in bias to ratings of behavior. This bias can affect performance ratings or ratings of potential employees in interviews. The drift may not always stem from the individual observer. Sometimes when many observers work together on the same job, they seem to drift away from the original definitions of the behavior, but as a group. Observer drift and contrast effects suggest that observers should be periodically retrained. Observer drift dissipates when there are frequent meetings to discuss methods. (Kaplan, 1997)

7.33 Expectancies

There is considerable inconsistency in the research of the effect of expectancy on behavioral observation. Some studies have shown that behavioral observers will observe the *behavior they expect*. On the other hand, some very thorough studies do not support an expectancy effect. Expectancy tends to cause bias in the behavior or observation when observers receive reinforcement for recording a particular behavior as opposed than when they do not.

The impact of expectancy is subtle. It probably has some minor biasing on the behavior observed. The finding that expectancy bias occurs significantly in some studies but not others is consistent with the notion that expectancy produces *a minor but potentially damaging affect*. To avoid the expectancy bias, observers should not be told what behavior to expect. (Kaplan, 1997)

7.34 Lying

Systematic study shows that most people do a remarkably poor job in detecting a liar. Not only are average people fooled by deception, so are police, judges and psychiatrists. When setting behavioural expectations of children, parents, teachers and workers in the plan of care, the treatment team leader should never take any statement of what happened for granted. Seek collaboration and examples of every critical observation. If the observation seems too good to be true, it probably is.

8.0 Item Analysis of the Indicators of Success

The plan of care starts with an assessment of the skills and accomplishments of the child. The content of the list of skills and accomplishments was drawn from the literature on emotional well being, resilience, attachment, social support and social skills. These are organized under a hierarchy of vision statements and life span outcomes.

8.1 Vision #1 Basic Functioning

The child will have social skills to function and grow within the protective structure of home and school.

There are three “life span outcomes” that are associated with this vision for the child:

- Happiness or contentment most of the time
- Ability to care for others
- Capacity for creative, productive work

According to an extensive literature review of the concept of “psychological well being” (Taylor & Brown, 1988) all of the great psychologists (such as Maslow and James) agree that these three outcomes define the most basic elements of a successfully functioning person. The indicators of success which flow from these outcomes represent the different ways that these abstract concepts have been measured in the research, according to Taylor and Brown..

8.2 Vision #2 Social Support

The child will have the support and acceptance of his peers. Community centres and neighbors will be positive towards the child, especially by including in their daily lives.

There are three “life span outcomes” that are associated with this vision for the child:

- good informal support network
- neighborhood supports
- ability to use available supports

These life span outcomes are essential to competence in society and family life. Cameron & Vanderwoerd (1997) reviewed the research supporting this self evident truth. The actual indicators are the behavioural evidence used by researchers to demonstrate that these outcomes existed.

For example, these are the indicators of success of people who acquire a good informal support network (Cameron, 1997 ,page 18)

“the child has available to him (her) when needed

- tangible or material resources
- people to do things with
- people who can give advice
- people that say “you are OK” and provide emotional support

8.3 Vision #3 Safe, Supportive and Accepting Care Givers

The child will live in a family that is an accepting, supportive and safe environment.

There are two “life span outcomes” that are associated with this vision for the child:

- **supported and supportive family**

The validity of the indicators for this outcome came from several sources,

1. is securely attached to someone in his life (Bowlby, 1988, Ainsworth, 1991, Fonagy, 1993)
2. has someone capable and motivated to advocate for him (discussions with group home staff)
3. his/her family members are coping well with their feelings about the youth. They
 - are not pessimistic about the future
 - are able to meet their needs as people
 - include the youth in their family events (Friedrich, 1983)
4. his/her parents have good behaviour management skills (August, 1995;
5. his/her parents have good social support in four areas: (Cohen, 1984)
 - tangible or material resources
 - people to do things with
 - people who can give advice
 - people that say you are OK and provide emotional support

All but item #2 has been validated through extensive research. Indeed, many of the threshold {b} tools measure these items. Within the context of the case planning model, the children, parents and workers assess these items (in terms of strength, OK, needs work, priority) strictly as a behavioural observation. However, we will validate the observation through statistical procedures with the instruments in risk screening, or through the addition of supplementary testing

- **parent's ability to cope with stress and conflict**

The validity of the items under this outcome is as follows:

His parents have normal levels of parenting stress in terms of:

- personal needs
- parent child interaction
- concern about future of child

The items listed above are taken from the Parenting Stress Index by Abidin. These items and the corresponding measurement instrument have proven to be excellent predictors of child abuse

(Burrell, 1994; Chan, 1994)

- there is good communication, mutual respect and support between parents
- parents display marital fidelity and commitment to the family

There are numerous references to the validity of the above two items both for predicting child abuse deaths (Kasim, 1995) and conduct disorder (Kagan & Kasdin, 1995). Several instruments exist to measure these items, including the FAM, the FACES and the FES. We did not employ these instruments in the risk screening protocol because the evidence of harm from the pathological state (marital strife, adultery, marriage breakdown) is very ambiguous.

Family dysfunction specifically marital conflict, divorce, problems in cohesiveness and parental irritability of the child, are much less strongly associated with new disorders arising during adolescence, but affect children under the age of 10 years. (Rutter, 1980) In fact, if you selected children from the total population where the parents are in conflict or divorcing or score poorly on psychometric scales of cohesiveness and responsiveness to children, the vast majority of these children would have no disorder or social problems. Two longitudinal studies by Hetherington & Camara and Wallerstein & Kelly (reviewed in Clapp, 1988) of children in divorcing families found that two thirds of children adjust successfully within two years of the divorce. Moreover, it is the persistence of problems in the area of marital conflict which causes disorder and child abuse, not episodic dysfunction. Some difficulties in marital conflict, cohesiveness, adaptability, control and responsiveness to children exist in every family on a continuum and there are only two conditions of family dysfunction which consistently produce disorder in children:

- [a] extremely poor cohesiveness, adaptability, parental control or responsiveness
- [b] very rigid who families who never move along this continuum (Clapp, 1988; Factor, 1990)

Nevertheless, family difficulties in marital conflict, communication, cohesiveness, adaptability etc. are very important clinically and are primary targets of intervention, especially for young offenders (Andrews, 1992) and in the treatment of child abuse (Factor, 1990). The presence of these qualities in a positive sense is very supportive to children and a worthwhile strength to celebrate and reinforce in treatment.

8.4 Vision #4 Health promotion and avoiding risk taking behavior

The child will be healthy and will avoid risk taking behavior

There are three life span outcomes associated with this vision statement. The concept of a “life span” outcome refers to a state of being that hopefully will endure throughout their entire life. Certainly if a child were OK in relation to all of the indicators under each life span outcome, they would have a secure hold on this outcome that would likely last much longer than their time in

care. The life span outcomes under this vision statement are priority areas of the federal government's ministry of Health and Welfare:

□ **good health related behavior**

Some of the behavioural indicators of success under "good healthy related behavior" included: "Involved in team sports", "has an attitude of ownership and awareness of physical self", "displays a positive healthy body image", "is physically active" and "is alert and aware"

□ **supportive social context**

Qualities in the neighbourhood and school also have a powerful independent effect on children's psychopathology and child maltreatment. (Guerra, 1995; Rutter, 1981; Rutter, 1980) These qualities include:

- schools with high turnover of teachers and students, uninviting atmosphere, sense of failure and uncaring attitudes
- apartment complexes with large semi public areas
- socially disintegrating neighbourhoods

The importance of the indicators under *a supportive social context* are evident from the child's point of view (see the research by Tremblay); however, by placing these indicators within the social context, where they exist, the workers are able to target their interventions more exactly.

One of the behavioural indicators of success under a supportive social context is:

the child's teachers are optimistic, caring and accepting,

A Montreal study, which followed children from grade one to age 14 years, found that early school failure was linked to later psychological distress and deviant attitudes in adolescence, although not delinquency per se. These negative results from early school failure may be a function of the breakdown in the emotional bond between the child and the school and in setting up peer rejection. (Tremblay, 1995) Early school failure, however, is not indicated by reading problems or lower IQ; rather it refers to the child's adjustment to kindergarten and the primary grades.

□ **positive consumer of health & social services**

Teaching our children how to act in a responsible and effective way in relation to our health and social service system is an important public priority.

One the most potent skills leading to responsible consumer behavior is: *future oriented skills, such as planning ahead, saving, anticipating problems, setting personal goals and persistence*

8.5 Vision #5 Life Skills

The child will make a successful transition to the next level of independence consistent with his stage in life.

There are five major life skills included under this vision with a variety of indicators of success related to each. The major life skills are:

- Taking Care of Personal Needs
- Communication Skills
- Ability to Access Public Transportation
- Money Handling Skill
- Time Management Skills

Many agencies are using the Daniel Hawthorne Memorial Institute life skills program with its own computer driven assessment. The indicators of success under this vision are very similar to the Hawthorne life skills program and many others.

8.6 Vision #6 Resiliency (ability to cope)

The child will have the strength to face adversity (resilience)

The language of the life span outcomes and the indicators of success themselves were copied directly from the literature on resilience and represent the qualities which defined the resilient child. (For a comprehensive literature review see Fonagy, 1993)

By the time, the worker gets to Vision #6, it will be apparent that many of the same indicators of success are repeated under different life span outcomes and visions. The redundancy is not a problem, since the instrument Indicators of Success **is not like** the risk screening instruments, where a global score is added up from the different items. This form is a criterion referenced test, in which the skills and accomplishments of the child and parents are assessed relative to the visions and outcomes of the service. In life - as demonstrated in research - the same set of skills (such as good problem solving ability) often supports many positive outcomes.

8.7 Vision #7 success at school or work

The child will be successful in School or Work

Vision #7 provides twenty one (21) standard indicators of success grouped under six key result areas that ensure positive results in school or work. This vision statement is on the most popular themes for the goals selected as priorities under the plan of care in OARTY agencies. The key result areas and related indicators require changes to the skills and attitudes of the children placed that are simply the most difficult treatment goals facing our children. The key result areas are listed in bold italics.

□ ***accepting the roles of teacher (or supervisor)***

The behavioral indicators of success are “follows teachers instructions”, “accepts guidance and criticism from teacher” and “accepts changes in teacher or task assignments”

□ ***positive work habits***

The indicators are “starts work on time”, “works steadily during the entire period”, “conforms to rules and regulations” and “pays attention to details and quality of work”.

□ ***Behaviour with peers in school & work***

The indicators are: “accepts assignments to group tasks”, “supportive of others in group task”, “responds to and initiates conversation with others”, “controls inappropriate expression of feeling, esp anger or aggression” and “joins group activities that are social or friendship in nature”.

□ ***Social Skills in the school setting***

The indicators are: “maintains satisfactory personal hygiene habits”, “arrives appropriately dressed for school” and “shows good judgement about horsing around or casual remarks”

□ ***manual dexterity skills***

The indicators are: “can accurately move body parts in relation to specific job” and “can manipulate small objects or tools with fingers”

□ ***Communication skills in the school or work setting***

The indicators are: “has clear speech, “is able to read and write at the level required for the task assignment”, “listens to others and comprehends what they say” and “initiates speech across the

full range of content required for the workplace (e.g. to alert others of unsafe situation, to carry out task assignments, to ask for assistance, to thank others for assistance and to support social interaction or friendships)

The indicators of success under the vision of success in school and the workplace were adapted from unpublished criteria of success for developmentally disabled teenagers in the transition from school to the workplace and sociometric studies of behaviour disorder and factors maintaining the problem (Erhardt, 1994; Caplan, 1992).

9.0 Threats to the Truth in Outcome Research

The fact that we give the children and parents standardized tests may lull us into a false sense of security. There are dozens of ways that even the best instruments will give false conclusions about whether the client is progressing and what is helping him to do so. The following examples have been highlighted in the literature. (Nurius, 1995; Kaplan, 1997)

9.1 regression from the mean

Parents or children who score very high at the start of a treatment process get better naturally with or without intervention because everything in nature always reverts back to the average state of affairs. Parents that are stressed to the very limit of endurance will experience less stress somehow over time. No-one has the energy to stay stressed to the limit forever.

Adolescents who display extreme aggressiveness or even murder, quickly revert to a less aggressive stance. The problem is that the aggressive teen is much more vulnerable to future displays of dangerousness if the social situation changes.

9.2 test data is not a substitute for real life outcomes

Sometimes the process of becoming less aggressive or less stressed is achieved by means of embracing an even worse outcome. The stressed out parent could alleviate his or her psychological state by simply giving up on the child or leaving the marriage altogether. The aggressive teenager could lower their presentation of fighting by escaping from their troubles through drug abuse, or by simply avoiding the challenges in life with which they are having difficulty. (Nurius, 1995)

9.3 all tests have a built in region of false positive and false negative

The Achenbach Child Behavior Checklist has been tested in longitudinal studies spanning six or more years with thousands of subjects. Sixty percent or more adolescents who score above 70T on the composite internalizing scale will present no symptoms one year later. At least 30% and in some studies 50% of teenagers who score above 70T on the composite externalizing scale will

present no symptoms one year later. The degree of false positives within the individual scales (withdrawn, anxious) is even higher.

There is also a well established false negative problem with the CBL. It under estimates the disturbance in {a} abused children, {b} teenagers who are developing into a sociopathic personality and {c} teenagers who are depressed and coping well socially.

These well documented problems with the Achenbach apply to all other tests. The Achenbach is actually one of the best instruments in relation to the error factor.

9.4 highly sensitive tests are vulnerable to short term, transitory events -

The Parenting Stress Index, for example, is very sensitive to the parent's inner psychological state of mind. This is good because high scores could indicate a serious depression or other mental health problem. The sensitivity of the PSI will also alert workers to the amount of psychological energy the parent has to apply to the presenting problem. However, the high degree of sensitivity in this instrument, means that transient pressure - such as a family visit by a needy, non-supportive relative - could push up the scores significantly. In this example, a "significant" change in scores, has very little significance clinically.

This example also illustrates how events that are totally outside of the therapeutic context can affect test results.

9.5 alpha inflation

This obscure title refers to the multiplication of error factors when using several psychometric tests for pre and post comparisons. This is also referred to in the literature as "experiment-wise error" (Peterson, 1995) Normally, the researcher selects a magnitude of change that has only a 5% chance of being a random event - in other words an error. If you have six tests or even six "scales", the probability that you will observe significant change due to random chance alone on any one of those scales is $6 \times 5\%$ or 30%. There are two solutions to the alpha inflation problem: {a} having a large number of subjects which decreases the value of alpha for the same t-score and {b} comparing change based on composite scores only (the technique used in the risk screening system) or creating your own composite from the multiple measures used. (Nurius, 1995; Peterson, 1995)

9.6 sleeper effect

Often the impact of a service will not be evident for many years. While evaluating the Headstart program in the USA, researchers were disappointed to see that poor children displayed the same level of academic success at the end of grade school, with or without Headstart. It wasn't until the end of High School that the true impact of Headstart became evident - Headstart children stayed in school longer.

9.7 The examiner test taker relationship

Both the behavior of the examiner and his or her relationship to the test taker can affect tests scores. Numerous studies have been done on the relationships between the examiners rapport and the child's performance on the IQ tests. Children who are familiar with the examiner produce an average house 4 IQ points higher; in those studies involving children from lower socioeconomic classes, familiarity accounted for about 7.6 IQ points. (Kaplan, 1997, chapter 6)

9.8 Expectancy effects

Sometimes the data can be affected by what an examiner expects to find. Robert Rosenthal and his colleagues at Harvard University conducted many experiments on such expectancy effects, often called the Rosenthal effect. After reviewing many studies, the conclusion is that an expectancy effect exists in some but not all situations. (Kaplan, 1997, chapter 6)

A variety of interpersonal and cognitive process variables have been shown to affect ones judgment of others. These biases may also affect test scoring. For example, researchers have demonstrated that test administrators are more likely to give credit for selected items to examinees that they like or perceived to be warm. Thus, examiners must remain aware that their relationships with examinees can affect their objectivity about certain types of tests.

Studies of expectancies in situations when the test administrator actually gives a score (e.g. the CGAS) rather than simply adding up and interpreting a test completed by the child or a parent, have yielded somewhat inconsistent results. Some have shown a significant effect, whereas others have not demonstrated an expectancy effect. In spite of these inconsistent results, you should pay careful attention to the potentially biasing effect of expectancy. (Peterson, 1995)

9.9 Effects of reinforcing responses

Because we know that reinforcement affects behavior, we should always administer tests under controlled conditions. An inconsistent use of feedback can damage the reliability and validity of test scores. Some of the most potent effects of reinforcement arise in the attitudinal studies. Repeated studies have demonstrated that the way an interviewer responds affects the content of response in interview testing situations. For example, respondents in a household survey were asked if they suffered from certain physical symptoms. For half the subjects, the interviewer gave an approving odd each time a symptom was reported. For the other half, the interviewer remained expressionless. The number of symptoms reported increased significantly with such approval. In a similar study, two symptoms that no one should report were added to the list: "are your intestines too long?" And "Do the ends of your hair itch?". More people reported the symptoms if they had been reinforced for reporting other symptoms than if they had not.

The Objective Stressors Checklist (threshold {b}) is vulnerable to reinforcement (either

positively or negatively) from the examiner.

Reinforcement's potency requires strict control over the use of feedback by test administrators. Because different test takers make different responses one cannot ensure that advantages due to reinforcement will be the same for all people. As a result most test manuals and interviewer guides insist that no feedback be given. Instruction for administering the objective stressors checklist require the examiner to follow the "script" fairly closely and to copy down the exact words of the child as he identifies the things he is worried about.

9.10 The absence of control or comparison groups

In order to isolate the specific effect of treatment on the changes in the test data or the plan of care results, evaluation theory (Peterson, 1995) requires that the researcher use at least two groups: an experimental group that gets the treatment and a control group (similar in all respects to the first, but who do not get the treatment). This is a critical strategy that has power to eliminate many of the sources of error listed above, including regression from the mean.

There are two techniques for coping for the absence of control groups: {a} OARTY will employ a single case design for selected target groups and research hypotheses (see section 10.0); {b} OARTY will set up comparison groups between the different models of treatment offered by the 70 member agencies. Children will be matched on risk screening profile and behavioural indicators of success on functioning levels. We can then compare the changes that result while controlling for resource ingredients (such as duration of treatment, intensity or staffing levels and treatment modality).

9.11 ensuring a truly representative clinical group

In order for group statistics to have any meaning or relevance to the field, the nature of the clinical group under the research must reflect the typical child going into treatment. Since, the study group is quite simply every child being admitted to the largest network of residential beds in the Province of Ontario, this methodological pitfall (so common among university based studies) has been fully solved. Secondly, the research will track the same child if he moves to another residential program providing further continuity.

9.12 the lack of random assignment

Random assignment to experimental and control groups is a standard method to ensure that the groups are similar before treatment. When random assignment is not possible, group equivalence must be demonstrated on every variable that could impact on the results, such as duration of symptomatology, severity level, age, sex, SES, etc. "Arguing with persuasive data that assignment, though nonrandom, was not related systematically to treatment relevant variables

and demonstrating group equivalence on a variety of subject characteristics can allow an investigator to cope efficiently with nonrandom assignment” (Peterson, 1995, page 156)

9.13 anchoring bias

definition: the excessive weighting of initial information about a client that subsequently serves as a template against which further information is judged.

There is a tendency with some referral sources to present the child in a way that is worse or more risky than he really is. Moreover, the tendency for all extreme conditions to revert back to an average level (9.1) amplifies this bias. (Nurius, 1995) The children, who will appear to have benefited least from intervention are in the middle range of severity. The solution to this bias is to use multiple measures and triangulation, both of which are feasible given the combination of the risk screening and plan of care model.

9.14 available heuristic bias

definition: refers to how easily certain information can be brought to mind

Workers or teachers remember the extreme behavior, but it is the typical everyday pattern which endures over time. Researchers have noticed that teachers take a much longer period of time to perceive changes in behavior as measured on behavior checklists, compared to parents. Also, school yard peers will continue to reject a child that acted as a bully - long after he ceased his bullying behavior.

This bias is often referred to as labeling theory. Please note: psychometric instruments are not immune to “invalid measurement” due to the effect of labeling on the respondent (Nurius, 1995)

9.15 covariance heuristics - bias

definition: the tendency to look for correlations that perhaps do not exist and then act upon them as causally related variables

This error occurs when we try to attribute the successful rehabilitation of an antisocial teen to our intervention (e.g. boot camps) without sufficient controls. The changes in the family or the child may have nothing to do with the intervention - or it may be due to some component of what you are doing. In some situations, the intervention is made up of helpful activities alongside other activities that actually reduce the positive results. In outcome studies of successful boot camps in the USA, it wasn't the “strict macho discipline” that produced positive results, but two other factors: getting the teens a job in the military after their time in camp and the out-patient support

post discharge.

The techniques to correct the later three sources of bias are: (Nurius, 1995)

- {1} establish a baseline on typical behavior/affect

The initial risk screening data meets this criteria.

- {2} employ the proper statistical procedures for computing the difference in means from time one to time two
- {3} structure the data collection to acquire the ability to check out correlation precisely ... record the **sequence** of events or observations

Psychometrically sound tests will not do you any good if the database around the testing is not comprehensive and accurate. This means keeping really good records of the **dates** that the children was referred, tested, admitted to the program, experienced other program changes and discharged from the program. The integrity of the OARTY outcome research is dependent upon and supported by the case management information system, which was implemented at the same time as the research.

10.0 Two Fundamental Designs to identify True Outcomes

There are two core designs that researchers have devised to avoid making incorrect conclusions:

- the single case design
- the group design

10.1 the single case design

The single case design selects from one to six subjects and follows their progress through treatment. The strategies used in single case designs to “check out assumptions” and eliminate error are as follows:

- {a} Single case designs will retest the child 10 to 15 times on the key outcome tests. The evidence of progress comes from a consistent pattern of positive movement on the test scores.
- {b} Single case designs also use at least ten different scales to “test” the clinical issues from many different angles.
- {c} Single case designs collect exquisite detail on the service history and on any other

“event” happening in the life of the child or his family that may affect the clinical outcome.

- {d} Single case designs often “experiment” with the intervention, such as withdrawing a specific component of the service, to see if the child reverts back to his old ways of behaving.

10.2 the Group design

The group design requires an absolute minimum of 15 matched subjects. The subjects making up the group must be similar on core socio demographic variables and service history. In most group designs, reported in the literature, the researcher would select one hundred or more subjects divided up into three or four “groups”

The special requirements of the group design are:

- {a} every member of the “group” receives a standardized battery of tests or questionnaires. The set of tests in the risk screening function is a typical set of tests for a group design.
- {b} The set of tests are administered before intervention and close to the termination of service. Ideally, you should have one or two extra tests between pre and post test periods.
- {c} Ideally, you should have a control group, which is matched to the “treatment group(s)” except that the controls do not receive the intervention under study. Indeed, it is through the mixing and matching of comparison groups that “experimentation” occurs.

One technique to substitute for the “no treatment control group” is to test children on the waiting list at the point of referral and then again at the point that service commences. As a result, you have two “pre” tests and one post test. Each child becomes in effect his own control. Ideally, the two “pre” tests should show little or no change compared to the post test.

For any single subject on any single test, there is a high probability of error. In the group design, the built in error factor is managed by averaging the pre-post change in scores over many subjects. *As long as the error factor is randomly distributed across subjects and tests*, the process of averaging the result will eliminate the error. The only problem with group designs is the danger that the error factor is systemic; in other words, the same error is occurring in a class of subjects or on one particular test. The classic systemic error factor is cultural bias. For example, several tests which measure family functioning (FAM, FACES, FES) run into serious

difficulties when you have a racially mixed group of subjects. A high score on “authoritative parenting” (fostering independence and encouraging the child to comply with parental rules by explaining the purpose and meaning of the rule) predicts social competence among white middle class children. Conversely a high score on “authoritarian parenting” (emphasizing compliance as a virtue and the use of punishment to obtain compliance) predicts social competence among African-American middle class girls.

Also, the **perception** of teachers or parents on tests measuring non-compliance, irritability or aggression is either exaggerated or minimized in some cross cultural combinations of the “tester and subject”. Finally, the objective reality of behavior problems is itself greatly influenced by teachers and students or parents and children who are mismatched in cultural terms.

In group designs, cultural bias is handled by averaging out the problem; however, this requires the researcher to check out the possibility that some factor is **interacting** with the testing process in a systemic way to produce the observed results.

10.3 accountability for the single subject

The rules for both the single case design and the group design have a devastating impact on agency plans to use this data for accountability purposes - in other words to ensure that each and every child is progressing and that children who are not progressing are identified and reviewed.

Psychometric testing is simply not accurate enough to become an efficient and effective agent of quality assurance at the case level.

Researchers in Britain’s child welfare system (see ONLAC) have studied this issue at length and their response was to devise a lengthy structured interview that they administer to every child in the system once a year in order to be assured that their needs are being met. Other techniques used to ensure clinical accountability include random audits by supervisors and exit interviews. These other approaches are much more efficient and accurate for purposes of clinical accountability. The limitation of these quality assurance techniques is that they are totally individualized and open ended which means that you have no common ground to sum the results and generalize the findings. However, these techniques do provide useful qualitative information about the impact of a service.

11.0 Hot Issues in the Practice of Testing

11.1 Professional Issues

There are three theoretical issues behind all testing protocols:

{a} all tests are grounded on a set of beliefs and theories about human functioning and tests are “no better than the theories and assumptions that underlie them” (Kaplan, 1997, page 609);

{b} there is no such thing as a bullet proof, rock solid, absolutely correct test. Even the IQ tests, upon which so much of a child’s plans in school are predicated has significant limitations in validity; we should never stop questioning and testing the validity of assumptions arising from testing;

{c} actuarial versus clinical prediction: should we rely on the test data to make predictions about the future behaviour and state of well being of a child or his parents or should we rely on the clinical judgement of a well trained professional? The research on this question is very conflicted. In general, actuarial models of prediction are more accurate, but only if they are methodologically sound. Using unreliable tests or making an invalid inference from a reasonably good test, is much worse (in terms of predicting human behaviour) than simply using thoughtful clinical judgement (Webster, 1994)

All three of these issues imply that great care must be taken when making critical decisions about a child based on his performance in various sociometric tests.

11.2 Moral Issues

Important moral issues in testing (Kaplan, 1997, page 611) are as follows:

{a} Human Rights

The American Psychological Association has declared that every human being has the following rights: *the right ...*

{a.1} *not to be tested*

{a.2} *to know test scores and interpretations*

{a.3} *to know the bases of any decision affecting their life*

{a.4} *to know who will have access to the data*

{a.5} *to confidentiality of test data*

{b} Labelling

Whenever, testing confirms a label that is feared or despised in public (such as “high risk” or “sex offender”) the individual may well be stigmatised for life; they may be subject to self fulfilling prophecies, thereby limiting his access to treatment. Finally, the label may reinforce a belief in the individual that he is not responsible or able to control his destiny, when he must learn to take responsibility for his life in order to get better.

The APA states that a person has a right not to be labelled and suggests that he should be described instead.

{c} Invasion of Privacy

A person’s privacy is invaded when personal information shared during or within the testing structure is used inappropriately. Professionals are bound in law (under the CFSA) to maintain confidentiality to reveal no more information about a person than is necessary to accomplish the purpose for which the testing is done. Moreover, professionals must inform the child or parent about the limits of confidentiality and that they have a right not to be tested.

11.3 Ethical Issues

There are two critical ethical issues of testing (Kaplan, 1997, page 613):

{a} Divided Loyalties

As professionals we must protect the welfare of our clients; but we work for institutions. A conflict arises when the welfare of the client is at odds with the institution. (For example, if a prospective foster parent is ruled unacceptable on the basis of her response on testing, does the professional side with the wishes of the applicant (and suppress the results) or the institution that employs them. The solution is that professionals must declare - *prior to testing* - where their loyalty lies and what the purpose of the testing is to the institution that is funding the test; secondly, the professional must describe the limits of confidentiality. To the institution, the professional should reveal only the minimum information needed. Unnecessary or irrelevant information remains confidential.

{b} responsibilities of test users and test constructors

Test users are responsible to know

- {b.1} the reason for using the test
- {b.2} the consequences to the client

- {b.3} the procedures necessary to maximise the test's effectiveness
- {b.4} the procedures necessary to minimise unfairness
- {b.5} the principles underlying the construction and supporting research of the test
- {b.6} the psychometric qualities of the test and the literature on its use
- {b.7} that interpretations based on the test are justified
- {b.8} that the test is properly used

It is the test constructor's responsibility to provide all of this information. Standards for the content of the test manual has been published by the APA. "Despite these guidelines, tests that do not meet specified standards continue to be published. A test user has no excuse for employing an inadequately documented instrument that has damaging consequences" (Kaplan, 1997, page 614)

11.4 Social Issues

Three social issues that have emerged in the literature are:

- {a} dehumanisation

Many people are very concerned about the proliferation of computerised testing, scoring and "cook book" interpretations. As clearly evident from the foregoing review of issues, the testing process is far from perfect and very challenging to a professional. The human element cannot be deducted from the process without increasing geometrically the error and potential damage to clients. The important principle of the OARTY system is that the specific instruments of risk screening were selected to ensure the wide involvement of staff and foster parents in the administration and scoring of tests.

- {b} usefulness of tests

The crucial social issue in testing is not whether tests are perfect but whether they are useful to society. The risks include the possible misuse of tests, which in turn may adversely affect the life of an individual or may discriminate systematically against a specific cultural group. The benefits include the potential for increased precision and fairness in the decision making process.

- {c} access to psychological testing services

The full set of psychological testing (including intelligence, projective, neurological and psychiatric testing) can be very costly; yet by denying access or providing insufficient funding for a child to have these tests, we may be making a life altering decision. Deeper testing may reveal dangers and problems that could be treated successfully if the need for intervention is recognised in time.

Consider what happened to Aboriginal people who suffered from Otitis Media, an inner ear disease. McShane (1988) found that 75% of Native children on US reserves suffered from Otitis Media. One of its consequences is mild to moderate hearing loss. Otitis Media is most prevalent between birth and the six or seven years of age during that the most critical developmental period for language. Other problems include delayed skin cognitive and the psycho linguistic development, delays in educational achievement, delays in grade placement, reading problems and emotional difficulties. School officials incorrectly diagnosed these Native children as developmentally disabled when in fact they were partially deaf. However, by treating the child as if her were “retarded” and giving him less challenging classes - instead of treating the hearing problem – “schools have effectively increased the incidence of *mental retardation* among Native children”. (McShane, 1988)

The use of risk screening instruments is *just the first step* in the assessment of a child; it opens the door to many possibilities that chart the future of a child. We are obliged to complete the assessment on a child - and to involve other professionals - especially if the risk data suggests that the child is facing a grim and limited future.

The risk screening data tells us how urgently we should respond to the child, but it will not tell how we should respond; as the risk increases, we are under increasing obligation to take the next steps in assessment and to advocate for the resources that will ensure a better future for the child.

Robert Fulton, M.S.W.

Friday, December 3, 1999

Bibliography

Achenbach, T.M. (1991), *Teacher Report Manual*, Vermont

Ainsworth, M. (1991), "Attachments and other affectional bonds across the life-cycle", in *Attachment Across the Life-Cycle*, Colin M. Parkes, Joan S. Hindes & Peter Marris (Eds), London, Routledge, pp 33-51

Andrews, D.A., Leisheid, A.W. & Hoge, R.D. (1992), *Profile, Classification and Treatment literature with Young Offenders: a social psychological approach*, Toronto, MCSS

August, G.J., Realmuto, G.M., Crosby, R.D. and MacDonald, A.W. (1995), "Community-Based Multiple-Gate Screening of Children at Risk for Conduct Disorder", *Journal of Abnormal Child Psychology*, Vol 23, No. 4, 521-544

Beiser, M. (1972), "Etiology of Mental Disorders: Sociocultural Aspects", in B.B. Wolman (ed), *Manual of Child Psychopathology*, New York: McGraw Hill, 150-189

Beitchman, J. H. & Raman, S. (1979), The Assessment of Childhood Psychopathology: The construction of a new self-report psychiatric rating scale for children", *Multivariate Experimental Clinical Research*, 4, 23-31

Beitchman, J. H. (1996), *Feelings, Attitudes and Behaviors Scale for Children (FAB-C)*, Technical Manual, Toronto: MultiHealth

Bickman, L., Karver, M., & Schut, J. (1997), "Clinician Reliability and Accuracy in Judging Appropriate Levels of Care", *Journal of Consulting and Clinical Psychology*, 65 (3), 515-520

Borduin, C., Mann, B., Cone, L., Hengeller, S., Fucci, B., Blaske, D., & Williams, R. (1995), "Multisystemic Treatment of Serious Juvenile Offenders: Long-Term Prevention of Criminality and Violence", *Journal of Clinical and Consulting Psychology*, 63, 4, 569-578

Bowlby, John (1988), *A secure base: clinical applications of attachment theory*, London, Routledge

Brown, M., Gordon, W. & Diller, L. (1984), "Rehabilitation Indicators" in *Functional Assessment in Rehabilitation*, Andrew S. Halpern & Marcus J. Fuhrer (eds), Paul Brooks Publishing, Baltimore, pp 187-203

Brandenberg, N., Friedman, R. & Silver, S. (1990), "The Epidemiology of Childhood Psychiatric Disorders: Prevalence Findings from Recent Studies", *Journal of the American Academy of Child & Adolescent Psychiatry*, 29 (1), 76-83

Burrell, B., Thompson, B., & Sexton, D., (1994) "Predicting Child Abuse Potential Across Family Types", *Child Abuse and Neglect*, 18 (12), 1039-1049

Cameron, G. & Vanderwoerd, J. (1997), *Protecting Children and Supporting Families*, Aldine de Gruyter, New York, 290 pages

Canetti, L., Bacher, E. Galili-Weisstub, E., Kaplan De-Nour, A. & Shalev, A. (1997), "Parental Bonding and Mental Health in Adolescence", *Adolescence*, 32 (126), Summer,381-394

Caplan, M., Weissberg, R., Grober, J.,Sivo, P., Grady, K., & Jacoby, C. (1992), "Social Competence Promotion with Inner-City and Suburban Young Adolescents: Effects on Social Adjustment and Alcohol Use", *Journal of Consulting and Clinical Psychology*, 60 (1) 56-63

Chan, Yuk Chung (1994), "Parenting Stress and Social Support of Mothers who Physically Abuse their Children in Hong Kong", *Child Abuse and Neglect*, 18 (3), 261-269

Clapp, G., 1988, *Child Study Research*, Lexington Books, Lexington, Mass

Cohen, S. Mermelstein, R. Kaamarck, T. & Hoberman, H. (1985), "Measuring the Functional Components of Social Support", in *Social Support: Theory, Research and Applications*, I. G. Sarason & B. r. Sarason (eds), pages 73-94

Coopersmith, S. (1967), *The Antecedents of Self-Esteem*, San Francisco: W.H. Freeman

Costello, E. & Angold, A. (1995), "Developmental Epidemiology", in *Developmental Psychopathology: theory and methods*, D. Cicchetti & D. Cohen (eds), John Wiley & Sons, New York, 23-56

Davidson, L. & Linnoila, M., eds, 1991, *Risk Factors for Youth Suicide*, Hemisphere Publishing Corp, New York

Dellario, D., Goldfield, E., Farkus, M. & Cohen, M. (1984), "Functional Assessment of Psychiatrically Disabled Adults" in *Functional Assessment in Rehabilitation*, Andrew S. Halpern & Marcus J. Fuhrer (eds), Paul Brooks Publishing, Baltimore, pp 239-252

Dishion, T.J. & Andrews,D.W. (1995), "Preventing Escalation in Problem Behaviors With high-Risk Young Adolescents: Immediate and 1 year Outcomes", *Journal of Clinical and Consulting Psychology*, 63, (4) 538-548

Dohrenwend, B., Levac, I., ShROUT, P., Schwartz, S., Naveh, G., Link, B., Skodol, A., & Stueve, A. (1992), ASocioeconomic Status and Psychiatric Disorders: The Causation-Selection Issue@, *Science*, 255, 946-951

- Erhardt, D. & Hinshaw, S. (1994), "Initial Sociometric Impressions of Attention-Deficit hyperactivity Disorder and Comparison Boys: Predictions from Social Behaviors and From nonbehavioral Variables", *Journal of Consulting and Clinical Psychology*, 62 (4), 833-842
- Factor, D. & Wolfe, D. (1990) "Parental Psychopathology and High Risk Children", in *Children at Risk*, Robert Ammerman and Michel Hersen, eds, New York, Plenum, 177-198
- Factor, D.C., Perry, A., & Freeman, N. (1990), "Brief Report: Stress, Social Support, and Respite Care Use in Families with Autistic Children", *Journal of Autism and Developmental Disorders*, 20 (1), 139-146
- Fergusson, D. M. & Horwood, L.J. (1995), "Early Disruptive Behavior, IQ. And Later School Achievement and Delinquent Behavior", *Journal of Abnormal Psychology*, 23 (2), 183-199
- Fonagy, P., Steele, M., Steele, H., Higgitt, A. & Target, M. (1993), "The Emanuel Miller Memorial Lecture 1992 The Theory and Practice of Resilience", *Journal of Child Psychology and Psychiatry*, 35(2), 231-257
- Friedrich, W., Greenberg, M. & Crnic, K. (1983), "A Short Form of the Questionnaire on Resources and Stress", *American Journal of Mental Deficiency*, 88 (1), 41-48
- Gagnon, C., Craig, W., Tremblay, R., Zhou, R., & Vitaro, F. (1995), "Kindergarten Predictors of Boys' Stable Behaviour Problems at the End of Elementary School", *Journal of Abnormal Child Psychology*, 23 (6), 751-766
- Gresham, F. & Elliott, S. (1990), *Social Skills Rating System, Manual*, American Guidance Services, Minnesota,
- Guerra, N., Huesmann, L., Tolan, P., Van Acker, R., & Eron, L., (1995), "Stressful Events and Individual Beliefs as Correlates of Economic Disadvantage and Aggression Among Urban Children", *Journal of Clinical and Consulting Psychology*, 63 (4), 518-528
- Guzder, J., Paris, J., Zelkowitz, P. & Marchessault, K. (1996), "Risk Factors for Borderline Pathology in Children", *Journal of the American Academy of Adolescent Psychiatry*, 35 (1), 26-33
- Harter, Susan (1985), *Manual for the Self-Perception Profile for Children*, Denver: University of Denver
- Hays, William (1965), *Statistics for Psychologists*, Holt, Rinehart & Winston, New York, 719 pages

- Holroyd, Jean (1983), *Questionnaire on Resources and Stress*
- Hudson, W. W. Nurius, P. S., Daley, J. G. & Newsome, R. D. (1986), *A short form scale to measure peer relations dysfunction*, Chicago: The Dorsey Press
- Kasim, M. & Cheah, I. (1995), "Childhood Deaths from Physical Abuse" , *Child Abuse and Neglect*, 19 (7), 847-854
- Kaplan, Robert M. & Saccuzzo, Dennis P. (1997), *Psychological Testing, Principles, Applications and Issues*, Fourth Edition, Brooks/Cole Publishing Company, CA, 724 pages
- Kazdin, Alan E and Jerome Kagan (1994), "Models of Dysfunction in Developmental Psychology", *Clinical Psychology: Science and Practice*, 1(1), 35-52
- Kazdin, Alan (1995), "Scope of Child and Adolescent Psychotherapy Research: Limited Samplings of Dysfunctions, Treatments and Client Characteristics", *Journal of Clinical Child Psychology*, 24 (2), 125-140
- Kovacs, M. (1992), *Children's Depression Inventory: Manual*, Toronto: Multi-Health
- Kruttschnitt, C., McLeod, J. & Dornfeld, M. (1994), "The Economic Environment of Child Abuse", *Social Problems*, 41 (2), 299-315
- Lessing, E., Oberlander, M. I. & Barbera, L. (1974), "Convergent of the IPAT Children's Personality Questionnaire and teachers' ratings of the adjustment of elementary school students, *Social Behaviour and Personality*, 2, 222-229
- Lochman, John E. (1995), "Screening Child Behavior Problems for Prevention Programs at School Entry", *Journal of Consulting and Clinical Psychology*, vol 63, No. 4, 549-559
- Loeber, R. (1990), "Development and Risk Factors of Juvenile Antisocial behavior and Delinquency", *Clinical Psychology Review*, 10, 1-41
- Loeber, R., & Dishion, T.J. (1987), "Antisocial and Delinquent Youths: Methods for their Early Identification", In J.D. Burchard & S. Burchard (eds), *Prevention of Delinquent Behavior* (pp 75-89), Beverly Hills, C.A., Sage
- Loeber, R. Dishion, T.J., & Patterson, G.R. (1984), "Multiple Gating: a Multistage Assessment Procedures for Identifying youths at Risk for Delinquency", *Journal of Research on Crime and Delinquency*, 21, 7-32
- Making Services Work for People, A new framework for children and people with disabilities*, Queen=s Printer for Ontario, 1997

- Mallinckroft, B. (1992), "Childhood Emotional Bonds With Parents, Development of Adult Social Competencies, and Availability of Social Support", *Journal of Clinical Psychology*, 39 (4), 453-461
- McShane, D. (1988), "American Indian Youth", *Journal of Adolescence*, 11,117-137
- Murphy, E., Brewin, C. & Silka, L. (1997), "The Assessment of parenting using the parental Bonding Instrument: two or three factors?", *Psychological medicine*, 27, 333-342
- Newman, F.L., Richard H. Hunter, R.H. & Irving, D. (1987) "Simple Measures of Progress and Outcome in the evaluation of Mental Health Services", *Evaluation and Program Planning*, 1987, vol 10, 209-218
- Nowicki, S. & Strickland, B. R. (1973), "A locus of control scale for Children", *Journal of Consulting and Clinical Psychology*, 40, 148-154
- Nurius, Paula and Walter W. Hudson, *Human Services, Practice, Evaluation and Computers*, Brooks Cole Publishing, Pacific Grove, Cal, USA 93950, 1995
- O' Donnell, J., Hawkins, J.D. & Abbott, R.D. (1995), "Predicting Serious Delinquency and Substance Use Among Aggressive Boys", *Journal of Clinical and Consulting Psychology*, 63, 529-537
- Offord, D., Boyle, M. & Ravine, Y. (1989), *Ontario Child Health Study, Children at Risk*, Toronto, Queen's Printer for Ontario
- Pancer, S. Mark (1986), "The Catalogue of Goals for Residential Treatment: An individualized outcome measure for residential Children's Programs", *The Canadian Journal of Program Evaluation*, 1(2), 69-79
- Gordon Parker, Hilary Tupling and L.B. Brown (1979), "A Parental Bonding Instrument", *The British Journal of Medical Psychology*, 52, 1-10
- Perry, A., Sarlo-McGarvey, N. & Factor, D.C. (1992), "Stress and Family Functioning in Parents of Girls with Rett Syndrome", *Journal of Autism and Developmental Disorders*, 22(2), 235-248
- Perry, A., Factor, D.C., Freeman, N, (1992), "The Effect of Parent Groups on Stress in Families of Children with Autism", *Children's Mental Health*, 5 (2) 18-23
- Petersen, Lizette and Debora Bell-Doran (1995), "Treatment Outcome Research in Child Psychology: Realistic Coping with the Ten Commandments of Methodology", *Journal of Clinical Child Psychology*, 24 (2), 149-162

- Piers, E. V. (1969), *Manual for the Piers-Harris Children's Self-Concept Scale*, Nashville, TN: Counsellor Recordings and Tests
- Rimmerman, A., & Duvdevany, I. (1995), "Coping Resources of Mothers of Integrated and Non-Integrated Pre-schoolers with Developmental Disabilities", *The British Journal of Developmental Disabilities*, Vol XLI, Part 1, No. 80, 42-47
- Rey, J. (1990), "Quality of Perceived Parenting in Oppositional and Conduct Disordered Adolescents", *Journal of the American Academy of Child & Adolescent Psychiatry*, 29 (3), 382-385
- Reynolds, J. R. (1980), "Concurrent Validity of the What I Think and Feel: The Revised Children's Manifest Anxiety Scale", *Journal of Consulting and Clinical Psychology*, 48, 774-775
- Rosenberg, M. (1963), *Society and the Adolescent Self Image*, New Jersey: Princeton University Press
- Rutter, Michael (1995), "Developmental Psychopathology: concepts and prospects", in *Developmental Psychopathology*, Dante Cicchetti & Donald Cohen (eds), J. Wiley, New York, 209-237
- Rutter, Michael and Seija Sandberg (1985), "Epidemiology of Child psychiatric Disorder: methodological Issues and Some Substantive Findings", *Child Psychiatry and Human Development*, 15(4), 209-233
- Rutter, Michael (1982), *Maternal Deprivation Re-assessed, second edition*, Penguin Books, Middlesex, UK, second edition
- Rutter, M., (1981), "The City and the Child", *American Journal of Orthopsychiatry*, 51(4), 610-625
- Rutter, Michael (1980), *Changing Youth in a Changing Society*, Harvard University Press, Cambridge, Mass
- (SA-45) *Symptom Assessment - 45, Questionnaire (1999)*, Technical Manual, MultiHealth, Toronto, 113 pages
- Sameroff, Arnold (1995), "General Systems Theories and Developmental Psychopathology", in *Developmental Psychopathology*, Dante Cicchetti & Donald Cohen (eds), J. Wiley, New York, 659-695
- Schopler, E., Reichler, R., & Renner, B. (1992), *The Childhood Autism Rating Scale (CARS)*, technical manual, Western Psychological Services, Los Angeles, CA. 15 pages

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

Sparrow, S., Balla, D. & Cicchetti, D. (1984), *Survey Form Manual, Vineland Adaptive Behavior Scales*, American Guidance Services, Minnesota, 299 pages

Speer, D. C. (1971), "Behaviour Problem Checklist (Peterson-Quay): Baseline data from parents of child guidance and non-clinical children", *Journal of Consulting and Clinical Psychology*, 36, 221-228

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

Sroufe, L. & Rutter, M (1984), "The Domain of Developmental Psychopathology", *Child Development*, 55, 17-29

Stark, K., Humphrey, L., Laurent, J., Livingston, R., & Christopher, J. (1993), "Cognitive, Behavioral and Family Factors in the Differentiation of Depressive and Anxiety Disorders During Childhood", *Journal of Consulting and Clinical Psychology*, 61 (5), 878-886

The Syracuse Community-Reference Curriculum Guide for students with moderate and severe disabilities, (1994), Alison Fors, Roberta Schnorr, Luanna Meyer, Linda Davern, Jim Black & Patrick Dempsey (editors), Paul Brooks Publishing, Baltimore

Taylor, S. & Brown, J. (1988), "Illusion and Well Being: A Social Psychological Perspective on Mental Health", *Psychological Bulletin*, 103 (2), 193-210

Tolan, P., Guerra, N.G., & Kendall, P.C. (1995a), "Introduction to Special Section: Prediction and Prevention of Antisocial Behavior in Children and Adolescents", *Journal of Clinical and Consulting Psychology*, 63 (4), 515-517

Tolan, P., Guerra, N.G., & Kendall, P.C. (1995b), "A Developmental-Ecological Perspective on Antisocial Behavior in Children and Adolescents - Towards a unified risk and intervention framework", *Journal of Clinical and Consulting Psychology*, 63 (4), 579-584

Tolan, P. & Thomas, P. (1995), "The Implications of Age of Onset for Delinquency Risk II: A Longitudinal Study", *Journal of Abnormal Child Psychology*, 23 (2), 157-181

Traglia, J., Pecors, P., Paddock, G. & Wilson, L. (1997), "Outcome-Oriented Case Planning in Family Foster Care", *Families in Society: The Journal of Contemporary Human Services*, September-October, 453-462

Tremblay, R.E., Masse, B., Perron, D., LeBlanc, M. (1992), "Early Disruptive Behavior, poor School Achievement, Delinquent behavior and Delinquent Personality: Longitudinal Analysis", *Journal of Consulting and Clinical Psychology*, 60 (1), 64-72

Vaden-Kiernan, N., Ialongo, N., Pearson, J. & Kellam, S. (1995), "Household Family Structure and Children's Aggressive behavior: A Longitudinal Study of Urban Elementary School Children", *Journal of Abnormal Child Psychology*, 23 (5), 553-568

Verhulst, F., Koot, H., & Van der Ende, J. (1994), "Differential Predictive Value of Parents' and Teachers' Reports of Children's Problems Behaviors: A longitudinal Study", *Journal of Abnormal Child Psychology*, 22 (5), 531-546

Webster, C.D., Harris, G., Rice, M. Cormier, C. & Quinswy, V. (1994), *The Violence Prediction Scheme*, Centre of Criminology, Toronto

Werner, E. (1989), "High Risk Children in Young Adulthood: A longitudinal study from birth to 32 years", *American Journal of Orthopsychiatry*, 59(1), 72-81

Werner, Emmy, E & Smith, Ruth, S. (1992), *Overcoming the Odds: high risk children from birth to adulthood*, Ithica, N.Y., Cornell University Press, 280 pages

Wheaton, Blair (1983), "Stress, Personal Coping Resources, and Psychiatric Symptoms", *Journal of Health and Social Behavior*, 24, (September), 208-229

Wheaton, Blair (1996), "The Domains and Boundaries of Stress Concepts", *Psychosocial Stress*, Howard B. Kaplan (ed), San Diego, Academia Press, 29-70