

A review of issues relevant to the creation of a  
measure of disability in children based on the World  
Health Organization's International Classification of  
Functioning and Disability (ICIDH-2)

Adrian Angold, MRCPsych.

E. Jane Costello, PhD.

October 12<sup>th</sup>, 2000

A brief summary of key concepts contained in the ICDH-2 . . . . .	4
Health condition vs. disability . . . . .	4
Dimensions of disability . . . . .	4
A fundamental problem for the measurement of disability in general . . . . .	5
Why has the DSM-IV included the presence of disability as a defining criterion for most child psychiatric disorders? . . . . .	6
Disability and harmful dysfunction . . . . .	7
The question of informant . . . . .	8
Some characteristics of disability in children and adolescents from a developmental perspective	9
Current measures of disability in children and adolescents . . . . .	11
Types of measures . . . . .	12
A note on terminology . . . . .	12
1. Global Measures of Disability . . . . .	13
1A. Children’s Global Assessment Scale (C-GAS): . . . . .	13
1B. Columbia Impairment Scale (CIS) . . . . .	14
2. Domain-Specific or Multidimensional Scales . . . . .	14
2A. Child Behavior Checklist (CBCL) Social Competence Scale . . . . .	14
2B. Teacher Report Form (TRF) . . . . .	15
2C. Social Adjustment Inventory for Children and Adolescents (SAICA) . . . . .	16
2D. Vineland Adaptive Behavior Scales . . . . .	16
2E. Vineland Screener . . . . .	17
2F. Child and Adolescent Functional Assessment Scale (CAFAS) . . . . .	17
2G. Behavioral and Emotional Rating Scale (BERS) . . . . .	18
2H. Instrumental and Social Competence Scale . . . . .	18
2I. Child Health and Illness Profile-Adolescent Edition (CHIP-AE) . . . . .	18
3. Measures of Symptom-Specific Impairment . . . . .	19
3A. Child and Adolescent Psychiatric Assessment (CAPA) Incapacity Ratings . . . . .	20
3B. NIMH Diagnostic Interview Schedule for Children. Version IV (DISC-IV) . . . . .	21

The Impact of Culture and Context on the Measurement of Disability .....	21
How many kinds of disability? .....	23
How well do current measures assess ICIDH-2 constructs? .....	23
How applicable are these measures to other types of illness? .....	23
Structured interview methods for collecting information on child and adolescent diagnosis and disability .....	24
Assessing disability using structured interviews .....	25
Discussion document on the measurement of disability in children and youth .....	26
1. Further research on measuring disability in children and youth .....	26
2. Future directions .....	27
A. Use of the disability construct in the assessment of need, utilization, outcomes, and costs. ....	27
B. Use in revisions of DSM and ICD .....	28
Conclusions about the measurement of childhood disability .....	28
Summary Table .....	30
References .....	32

## A BRIEF SUMMARY OF KEY CONCEPTS CONTAINED IN THE ICIDH-2

In this section we outline some key concepts employed in the International Classification of Functioning and Disability, to set the stage for later comparisons with other approaches.

### HEALTH CONDITION VS. DISABILITY

A key conceptual distinction is made between “health conditions” (often referred to as “disorders” or “diseases”). Health conditions are broadly defined, but for practical purposes can be regarded as including any conditions classified in the ICD-10.

Disabilities are negative functional outcomes *resulting* from health conditions, involving significant deviation from or loss of “normal” or “expected” function.

### DIMENSIONS OF DISABILITY

Disability may be observed in any of three “dimensions” or “areas of functioning”:

**Body functions and structure**, with psychological functions being included as body functions. The ICIDH-2 Beta-2 draft invokes the concept of “impairment” in relation body functions and structure (pp16-17), noting that impairments are “problems in body function or structure as a significant deviation or loss” (p16). Point (12) on p17, however, expands this definition somewhat, by noting that there are four possible types of impairment:

1. Loss or lack.
2. Reduction.
3. Addition or excess.
4. Deviation.

Impairments are also defined in terms of population norms (p6, point (5)).

**Activity** is the performance of a task or action by an individual (p18). Activity limitations are difficulties an individual may have in the performance of activities. Such limitations must be actually observable, and are rated regardless of capacity or aptitude (p18 point (4)). Several additions and qualifiers to the concept of activity limitation are also provided:

Activities involve the “integrated use of body functions” at the individual level (p18 point (1)).

Activity limitations involve qualitative or quantitative alteration in the way an activity is performed in relation to generally accepted local population standards (p18 point (5)).

The successful use of assistive devices may reduce or eliminate an activity limitation, though an underlying impairment (of body function or structure) is still present.

**Participation** refers to an individual’s involvement in life situations. Participation restrictions are problems an may have in the manner or extent of involvement in life situations. The

participation dimension differs from the activity dimension in being manifest at the social level, rather than the level of the individual considered regardless of society (p20 points (5) and (6)). Restrictions in participation are, therefore critically dependent on factors external to the individual. Participation restriction refers to “the discordance between the observed participation and the expected participation of an individual without a similar health condition” (page 20 point (7)).

#### **A FUNDAMENTAL PROBLEM FOR THE MEASUREMENT OF DISABILITY IN GENERAL**

The set of distinctions embodied in ICIDH-2 is clearly an enormous step forward in clarity and utility when compared with the first ICIDH. Its very clarity, however, reveals a fundamental difficulty that lies at the root of developing a coherent classification of disease and disability that will allow standardized measurement of the latter across different diseases. The problem is that the levels of disability that *can* be measured depend on the level at which a disease is diagnosed. Take, for instance, phenylketonuria - a disease characterized at the molecular level by lack of phenylalanine hydroxylase. This disease can properly be regarded as producing impairments in body structure and functioning (at the neurological, psychological and physiological levels). Even elevation of blood phenylalanine itself can be regarded as being a functional impairment (b540 - impairment of general metabolic functions), because the disease can be defined in terms of the *cause* of that elevation. Thus, functional impairment, activity limitation and participation restriction can all be measured for phenylketonuria. If we consider a health condition like diabetes, however, we have a health condition that must be diagnosed at the level of functional impairment (elevated blood sugar and inadequate insulin response). Many health conditions can only be diagnosed at this level (including the vast majority of psychiatric disorders), which makes the assessment of functional impairment problematic for such health conditions, because the identification of the health condition is dependent upon the presence of functional impairment. Such conditions violate the conceptual separation between health conditions and disabilities.

This is not to criticize the ICIDH-2 for maintaining such a distinction. On the contrary, maintaining the distinction is helpful, because it places the nosology on a footing that will be increasingly productive as we learn more about the causes and consequences of health conditions. It does mean, however, that we cannot expect to be able to develop very useful general measures of *impairment* in the near future, because so many health conditions are currently only diagnosable in terms of the impairments they cause. Thus it is not surprising that the WHO DAS-II questionnaire, which is intended to be a generic measure of adult disabilities, deals only with activity limitations and participation restrictions.

In a sense, we have to infer the presence of the health condition from the presence of impairments. We can still maintain the separation between health conditions “inferred” from impairments of function and structure and their resultant *activity limitations* and *participation restrictions*, because, even in psychiatry, neither activity limitations nor participation restrictions are *logically* necessary requirements for making most diagnoses (as Spitzer & Wakefield, 1999 amply demonstrate in their critique of DSM-IV’s practice in this regard). Indeed, several empirical studies have indicated that symptom and impairment measures tap meaningfully different dimensions of dysfunction, which independently predict a variety of outcomes, including service use and prognosis (Shaffer *et al.*, 1983; Verhulst *et al.*, 1993; Green *et al.*, 1994; Costello *et al.*, 1996c; Hodges & Wong, 1997b; Lyons *et al.*, 1997; McArdle & Gillett, 1997; Settertobulte & P.,

1997; Hodges *et al.*, 1998; Lavigne *et al.*, 1998; Costello *et al.*, 1999). Maintaining a distinction between diagnosis and disability also enables us to investigate the interplay between the two (Hodges *et al.*, 1999; Ezpeleta *et al.*, In press) over time, and from a developmental perspective it is far from obvious that we should expect to see unidirectional causal arrows from disorder to disability.

#### **WHY HAS THE DSM-IV INCLUDED THE PRESENCE OF DISABILITY AS A DEFINING CRITERION FOR MOST CHILD PSYCHIATRIC DISORDERS?**

In the case of child psychiatric disorders, the inclusion of a disability requirement in many DSM diagnoses has usually been driven by practical concerns, rather than by any deep theoretical or empirically-based reasoning. General population studies using lay interviewers and respondent-based interviews like the Diagnostic Interview Schedule for Children or the Kiddie Schedule for Affective Disorders and Schizophrenia, Epidemiologic version, have produced enormously high rates of disorder in children and adolescents when no disability criteria were employed. To get the population prevalence rates of psychiatric disorder down to around 20% it was necessary to exclude many children who appeared to meet the *symptom* criteria for DSM disorders. However, there is plenty of reason to doubt that this is an inherent characteristic of child psychiatric disorders. Rather, there is every reason to suppose that it is a product of the characteristics of these particular interviews, and the diagnostic algorithms applied to them, because some other interviewer-based interviews (also conducted by lay interviewers) do not produce inflated rates of psychiatric disorder when employing pure symptom criteria, without disability requirements. However, we must face the fact that current child psychiatric measures of the type likely to be popular for cross-national studies do appear to have this problem.

We should also consider an important “theoretical” position supporting the requirement that impairment be present for a psychiatric diagnosis to be made. Many of the symptoms of child psychiatric disorders (shyness, fears, and disobedience for instance) are very common when they occur at low frequency. In the absence of good population norms for the range of frequency with which children have such symptoms, or when they are assessed simply by asking whether they “often” occur, it is impossible to apply proper cut points to the frequency distributions to divide the “normal” from the “symptomatic” children. The use of a disability criterion provides an alternative method for deciding who is “symptomatic”. Of course, this simply shifts the cut-point problem from the symptoms to the measure of disability, and this has presented many practical difficulties. This solution also takes a “principled” stand against the separation of diagnosis and disability, and, therefore, directly contradicts the WHO’s position in regard to disability.

The DSM’s approach, of including impairment as a diagnostic criterion, also leads to some diagnostic oddities. The clearest example concerns conduct disorder (CD). Consider a child who often lies, steals and bullies other children, and who is so successful in these antisocial pursuits that he remains undetected by adults. The child meets the symptom criteria for CD, but does not meet the disability criterion. If he is also pleased with his progress as a budding psychopath, he also fails to meet the distress criterion (no matter how distressed his victims might be). Suddenly one day he is caught by a teacher “eliciting” money from one of his victims, and excluded from school as a punishment. Now he has a disability (a participation restriction), and meets diagnostic criteria, but his health condition (CD) has not changed in any way. Where there is a need to correct apparent

over-diagnosis resulting from limitations in our symptom measures, what is really needed is better information about how to define and measure the symptoms and signs of the disorder.

Including disability in diagnostic criteria also tends to inhibit study of the relationship between diagnosis and disability. While such research is certainly possible, we must recognize that there is a counter-pressure exerted by the official nosology to “stick to the rules”, and that comparisons between studies that have “stuck to the rules” and those that have not are always problematic. From a developmental perspective, in particular, we do not expect disorders suddenly to spring, fully developed, into existence. Rather we are interested in describing (and hopefully preventing) the emergence of fully-fledged and disabling disorders. Requiring disability to be present if a diagnosis is to be made distracts attention from early manifestations of disorder. To draw an analogy with physical medicine, it is as though we refused to recognize that children had asthma until it was sufficiently severe to keep them from regular school attendance.

In summary, whereas child psychiatry still relies heavily on *impairments in functions and structure* to make a diagnosis (as does much of medicine), the current DSM-IV custom of including *activity limitation or participation restriction* as a requirement for most child psychiatric diagnoses, rather than as separate dimensions, has arisen for reasons that have little to do with the nature of child psychopathology, but much more to do with technical problems with the measurement of symptoms using certain types of diagnostic interview. The ICD has largely avoided this confusion, and offers an approach that is better suited to the study of the development of disorders.

#### **DISABILITY AND HARMFUL DYSFUNCTION**

As several people have noted (Wakefield, 1997; Spitzer & Wakefield, 1999), the DSM-IV is inconsistent in its requirements in relation to disability (“impairment”) from one part of the manual to another and from disorder to disorder, and its approach raises a host of conceptual, measurement and methodological problems. For instance, several of the criterion sets for childhood disorders make no mention of distress, disability or impairment (Rett’s syndrome, pica, rumination, feeding disorder of infancy, encopresis, reactive attachment disorder, and 3 disorders not-otherwise specified (NOS) - communication disorders NOS, attention deficit hyperactivity disorder (ADHD) NOS, and tic disorder NOS).

Wakefield has provoked debate in consistently presenting his conceptualization of psychiatric disorders (and medical disorders in general) as “harmful dysfunctions” in opposition to the DSM system. However, his definition shares with the DSM-IV the requirement that the existence of a disorder demands the presence of impairment at the level of the functioning of the whole organism. For instance, he states (Wakefield, 1997 p279) “Many things go wrong with various mental and physical mechanisms that do not deserve to be called disorders because they do not have sufficiently negative implications for the individual’s overall well-being.... The difference between dysfunctions that can be classified as disorders and dysfunctions that cannot be classified as disorders thus lies in whether the dysfunction causes real harm to the person.”

This approach, though based upon a set of theoretical principles about the nature of disease, and more consistently applied than is the case in DSM-IV, is equally susceptible to the critique we set out of the DSM-IV approach.

## THE QUESTION OF INFORMANT

One reason why respondent-based interviews may produce very high rates of disorder when used with children, although they produce lower rates with adults, is that in child and adolescent psychiatric research it is usual to collect data from both parent and child, and sometimes from other informants, such as teachers, in addition. Thus, each individual has two or more chances to be rated positively on a symptom. Most adult studies are based solely on self-reports, so each individual has only one chance to be positively rated. Given the low rates of agreement between informants in respect of a child's psychiatric symptomatology, it is to be expected that the use of multiple informants will substantially increase apparent rates of disorder. We can also expect that rates of disability in children will appear higher than they are in adults when multiple informants are used. This consideration leads naturally to the question of who should be the subject of interviews or questionnaires about disability in children and adolescents.

Until the 1980s, agreement between child and parent reports of symptomatology was widely regarded as being a test of the *validity* of *child* reports (Rutter & Graham, 1968; Herjanic *et al.*, 1975). However, subsequent research using all sorts of measures soon showed that only low levels of agreement among informants (correlation coefficients around 0.3 for agreement among children, parents and teachers) could be expected (Reich *et al.*, 1982; Stanger & Lewis, 1993). It is now considered that low levels of agreement amongst different informants about the child's clinical state are to be expected and do not invalidate the reports of any of them. Rather, each key informant (typically, child, parent and teacher) is seen as presenting a particular view of the child's problems. For instance, since teachers observe the child at school, it is not surprising that they may identify problems at school of which the parents are not aware. Similarly, since neither teachers nor parents spend much time observing children in the school yard, the child may be the only source of information about social difficulties encountered there. Indeed, it is precisely because agreement among informants is low that multiple informants are needed. Were agreement very high, collecting information from more than one informant would be redundant.

The problem is that disagreement among informants means that one has to decide how to weight the information from each informant in arriving at a diagnosis. Since it is uncommon for informants to invent fictitious symptoms, the simple rule of regarding a symptom as being present if any informant reports it usually suffices well enough. When symptoms are combined to make diagnoses, the usual procedure is to "ignore" the source, and to add up all positive symptoms from any source. Thus, a diagnosis of a major depressive episode (which requires the presence of at least five symptoms) might be made on the basis of three relevant symptoms being reported by the child (say depressed mood, anhedonia and excessive guilt), with two other relevant symptoms (perhaps sleep and appetite disturbances) being reported by the parent. Though some interview developers have recommended "reconciliation" discussions involving the interviewer, the parent and the child to clear up discrepancies between their reports, such a discussions are problematic in several ways. Firstly, to achieve their purpose, one informant must modify his or her story, but that means admitting being wrong, or at least uninformed. Secondly, it offers a chance for family members to become engaged in arguments with one another - an activity that certainly does not help the clinical or research process. Thirdly, the knowledge that such a discussion will occur could cause informants (e.g. drug-using adolescents) to withhold important information that they did not wish other informants (such as their parents) to hear about. Finally, in most research applications, one wishes to assure informants that what they say will not be revealed to anyone else, in which case



a “reconciliation” interview is ruled out.

Little attention has been paid to the specific issue of how much agreement there is among informants in relation to measures of disability, but there is no reason to suppose that disagreement will not be as prevalent there as it is in relation to symptomatology. There is reason to suppose, therefore, that disability should ideally be measured using multiple informants. The key informants are:

5. The target child.
6. A parent or guardian or residential caretaker.
7. A teacher or employer.
8. A peer or friend.

When a minor is both attending school and employed, there is reason to want to collect information from both teachers and employers.

There are ways to avoid the problem of generating unrealistically high numbers of diagnoses (and, by extension, of cases of disability). Interviews that use an experienced clinician, or a glossary written by experienced clinicians, to define precisely what level of symptomatology or distress constitutes a clinical level of severity, do not generate unreasonably high rates of diagnosis even when multiple informants are interviewed. The same principle can be applied to disability.

#### **SOME CHARACTERISTICS OF DISABILITY IN CHILDREN AND ADOLESCENTS FROM A DEVELOPMENTAL PERSPECTIVE**

While complete agreement about the definition of development is still a long way off, the following definition from Nagel (1957) serves our purposes here well enough:

*The concept of development involves two essential components: the notion of a system possessing a definite structure and a definite set of pre-existing capacities; and the notion of a sequential set of changes in the system, yielding relatively permanent but novel increments not only in its structure but in its modes of operation. (p.15)*

By “structure” Nagel means the way an organismic function or activity is organized, he is not referring only to anatomical structures. Thus the “structure” of speech involves all the anatomical, physiological, and psychological components by which speech is produced. The key point here is that the organism starts with some structure, which is then built upon to yield new abilities or functions. Abnormal development (for instance, the development of a disability) may occur in several ways, which we will illustrate using the achievement of a normal adult gait as an example:

1. The starting structure from which development proceeds may be abnormal, as when a child has a congenital limb defect.
2. There may be a developmental *delay*, where normal walking is achieved, but more slowly than is usual.
3. There may be termination of the developmental sequence before it is complete - for instance, becoming able to crawl, but not to walk.
4. There may be *regression* from successful achievement of the appropriate end-point of the developmental sequence, as might result from an accident that rendered an adult hemiplegic.

5. There may be a *qualitative deviation* from the normal developmental path resulting in an abnormal end-point, as might result from nutritional deficits that had caused severe rickets.
6. There may be a *quantitative deviation*, as when a person walks with a normal gait, but can walk only very short distances because of cardiac problems.

Each of these possibilities is associated with particular issues for the measurement of disability:

1. The degree of disability may increase over time, as the functions of the developmental pathway fail to develop in normal ways. For instance, a legless newborn may function quite adequately compared with normal newborns. Disability will only be apparent later when the locomotor functions of the legs would have been expected to have emerged.
2. The observation of delay requires the existence of developmental norms for the emergence of particular functions, activities and participations.
3. The end point of the developmental sequence must be identifiable. If comparisons are to be made across individuals or populations then it must be reasonable to assume the existence of similar (or equivalent) developmental endpoints across those individuals or populations. For instance, an activity limitation in reading could not result from a structural problem with the legs in populations unexposed to literacy training.
4. The identification of regression requires that norms for the achievement of the final function of the developmental sequence and the activities and participations that depend upon it be available.
5. The identification of qualitative deviation similarly requires that the qualitative limits of normal performance of a function, activity or participation can be identified. This is relatively straightforward for an activity such as walking. However, we lack developmental norms for many other functions, such as the balance between aggressive behavior and prosocial behavior. Neither can we necessarily suppose that the same norms will be appropriate from one culture to another.
6. A similar problem applies to the quantification of normal levels of activity or participation. How much walking "should" a person be able to do? How much unhappiness does it take to constitute a disability in the pursuit of happiness?

The measurement of disability is, therefore, dependent upon "expert knowledge" of the function, activity or participation that may be affected by a health condition. In essence there are four ways to tap such "expert knowledge" in a disability measure:

1. Explicitly define an "adequate" achievement. For instance, item S1 of the WHO DAS-II asks "how much difficulty do you have in *standing for long periods such as 30 minutes*". The questionnaire implies that one should have no difficulty standing for 30 minutes. "30 minutes" serves as an explicit definition of a "long period".
2. Require reference to community norms, of which the respondent is expected to be aware. For instance question S4 of the WHO DAS-II asks "How much of a problem did you have joining in community activities.....*in the same way as anyone else can?*"
3. Require reference to the respondent's own internal norms. Question S2 of the WHO DAS-II asks "How much difficulty did you have in taking care of your household responsibilities?"

Here no comparison is offered with any standard external to the respondent.

4. Use an external expert judge or expert system to determine from descriptions of function ability, activities and participation whether a disability is present. "Clinical judgment" relies upon the assumed (and unstandardized) knowledge of clinicians to generate such judgments. However, it is also possible to make the rules for making such judgments explicit and standardized (as in an interviewer-based interview).

It will also be apparent that the first three strategies are dependent upon the respondent making a link between the presence of a health condition and any disabilities attendant upon it. When considering parents and teachers as informants, we must ask how much they know about the health conditions of their children. For instance, in a large longitudinal community study of the development of mental illness (the Great Smoky Mountains Study) (Angold *et al.*, 1998) the majority of parents of children with a psychiatric diagnosis or disability stated that their child did not have a problem or need help, even though they themselves had reported the symptoms on which the diagnoses were based. Children themselves were even less likely to state that they had a problem or needed help. We must ask to what extent it is reasonable to believe that the available informants are capable of making appropriate connections between health conditions and disabilities. There may be reason to doubt whether a statement like the preamble to the WHO DAS-II ("by health condition I mean diseases or illnesses, other health problems that may be short or long lasting, injuries, mental or emotional problems and problems with alcohol or drugs") will be sufficient to orient informants about children to the appropriate considerations.

Option 4 above offers an alternative approach that is sometimes supposed only to be feasible when clinician interviewers are available. However, this impression is mistaken. The issue here is not clinical training, but adequate procedures for implementing the appropriate decision rule. These decision rules need to be written by clinical experts, but they can be made available to interviewers in written form, or implemented as computer algorithms after the necessary data have been collected by an interviewer.

#### **CURRENT MEASURES OF DISABILITY IN CHILDREN AND ADOLESCENTS**

The measurement of functional impairment or disability as applied to children is in its early stages. None of the currently available measures was designed specifically to make the conceptual distinctions made in ICIDH-2 between impaired functioning, activity limitation, and participation restriction. In particular, most of these scales and interviews contain items that are symptoms and signs of health conditions (as defined in the ICD-10). In particular, most of these measures include ratings of emotional and behavioral deviations that are referred to in the ICD-10 criteria for psychiatric disorders. For example, an affirmative answer to the question "has your child seriously hurt or beaten someone up?" on the Child and Adolescent Functional Assessment Scale (CAFAS) carries with it a maximal level "impairment" rating. However, this item obviously pertains to the ICD-10 criteria for conduct disorder. The Child Health and Illness Profile (CHIP) includes items relating to "physical discomfort" which, in the ICIDH-2 system is regarded as being a potential *cause* of disability, rather than being a disability in itself. Many of the measures covered here relate more closely to what we might call "overall illness severity" than to the ICIDH-2 construct of disability. Of course, all of these measures were developed before the ICIDH-2 was available, so it is not surprising that they do not directly implement its concepts. On the other hand, all the

measures discussed here contain at least a subset of items that are congruent with the ICIDH-2 conceptualization of disability.

Different branches of medicine, psychology, and psychiatry have developed measures to tap different areas of disability (usually referred to as “functioning”, or “functional impairment”), many of them somewhat overlapping. For example, a 1998 review of measurement systems for child and adolescent mental health (Bickman *et al.*, 1998b) includes 11 measures of Coping Skills, 44 measures of Family Functioning (including the child’s functioning within the family), 10 of Functioning Competence, 6 of Functioning Impairment, 19 of symptoms, which would include many items relevant to ICIDH-2’s impaired functions and structures, and 17 multidimensional measures. The present review is restricted to measures from the Functional Competence, Functional Impairment, and Multidimensional areas that have adequate psychometric data. We include interviews as well as the paper-and-pencil measures reviewed by Bickman et al., and include new measures since the publication of their review. We have not included instruments designed purely to assess psychiatric symptoms, although as noted earlier these often overlap with the ICIDH-2’s *impairments of functions and structures*. Nor have we included measures of the individual’s satisfaction with current functioning, or self-esteem, or instruments designed specifically as measures of intelligence or school attainment. While the latter are, of course, measures of functioning, and have their role in the clinical assessment of disability, they were not designed to measure functional disability in the general way that the other measures described here were, and deserve separate consideration as part of a clinical assessment of functioning.

In this section we briefly review measures on which there are basic psychometric data in the form of test-retest reliability and construct validity. We have included all the measures on which test-retest data are available, and a selection of those on which no test-retest data are available but which despite this weakness appear to have useful qualities, such as good predictive validity. This means that some promising new measures are not reviewed here because they lack any relevant psychometric data. It should be noted, however, that none of the measures omitted from this review was designed specifically to fit the ICIDH-2 framework.

## TYPES OF MEASURES

Measures of functional impairment can be divided into four types:

1. Global disability measures;
2. Domain-specific or multidimensional disability measures;
3. Diagnosis-specific disability measures;
4. Symptom-specific disability measures.

## A note on terminology

Most measures define themselves as assessing “functioning” or “functional impairment”. By this they often refer to areas that in the ICIDH-2 would fall under the heading of *activity limitation* or *participation restriction*. Alternatively, “impairment” may be used to refer to any or all three ICIDH-2 categories of disability. In this review we have continued to use the term “impairment” in this global or non-specific way when the authors of the measures do so. Otherwise we use the general

term “disability” for the entire construct to which the ICIDH-2's threefold classification refers. We have tried, however, to note the extent to which a measure evaluates each of the three ICIDH-2 levels of disability.

## 1. Global Measures of Disability

Global measures of impairment enable the interviewer to rate a child on a single scale of functioning, from excellent to poor. Descriptions of typical functioning at various points along the scale are provided to aid the decision-maker. The advantage of a global impairment measure is that they are very short and easy to administer with little training, by lay interviewers. However, these scales do not provide for a separate evaluation in different areas of functioning, and are therefore of limited usefulness for treatment or prevention planning. Also, they are dependent on prior knowledge of the child's symptoms and level of functioning; they summarize the interviewer's pre-existing information, but do not replace instruments that collect that information.

### 1A. Children's Global Assessment Scale (C-GAS): (Shaffer *et al.*, 1983):

The respondent makes a single overall judgment, on a scale of 1 (most impaired) to 100 (highest level of functioning), of the child's functioning in the home, community, and school, and with friends. Each ten-point segment is anchored by brief descriptions of a typical level of functioning for a child with a score in that range. The time frame can be specified by the researcher, and the relevant level of functioning can be defined; e.g., the highest, or lowest, or average level of functioning within a given period. The scale was originally designed for clinicians' use, on the lines of the Global Assessment Scale for adults (Endicott *et al.*, 1976). Recently, a version for lay interviewers has been developed, similar to the clinician version but with simplified descriptors. The C-GAS is a single overall rating that is not keyed to any specific questions. It depends on the decision-maker's knowledge of the child. It was originally designed to be completed by interviewers who had just conducted a full-scale psychiatric diagnostic interview. Versions for use by non-clinicians have been developed (Hodges *et al.*, 1992; Setterberg *et al.*, 1992). It has also been completed by parents following a full-scale psychiatric diagnostic interview. Its use in relation to non-psychiatric conditions has not been examined, and the key descriptions of levels of functioning clearly focus on issues particularly pertinent to psychiatric conditions. On the other hand, when used by clinicians who know the child well, the scale takes only a few seconds to administer. The inter-rater and test retest reliability, as well as the concurrent and construct validity of both the clinician and the lay interviewer C-GAS has been reported (Shaffer *et al.*, 1983; Bird *et al.*, 1987; Bird *et al.*, 1990) in both Anglo and Hispanic samples. As can be seen in Table 1 the reliability and validity results are generally acceptable, but lower for the children's than the parents' versions. An advantage of using the scale is that training takes less than one day for lay interviewers, and less time for clinicians. Disadvantages include: (1) The descriptive examples confound psychiatric symptoms with indices of functioning. Consider, for instance, the following definition for a score of 70-61: “Some difficulty in a single area, but generally functioning pretty well, (e.g. sporadic or isolated antisocial acts, such as occasionally playing hooky or petty theft; consistent minor difficulties with school work, mood changes of brief duration; fears and anxieties which do not lead to gross avoidance behavior; self-doubts). Has some meaningful interpersonal relationships. Most people who do not know the child well would not consider him/her deviant

but those who do know him/her well might express concern." here symptoms like *fears* and *anxieties* ("impaired functions" in ICID-2 terminology) are confused with activity limitations (*difficulties with school work*). (2) The C-GAS rating is dependent on the interviewer's judgement of what constitutes normal functioning, so it may be less useful when used with children of diverse cultural backgrounds. While a score of 60 or below has been widely used in studies in the United States to define functional impairment, preliminary work to identify an appropriate cut point is recommended if the C-GAS is to be used in cultures or subcultures other than those in which it was developed.

### **1B. Columbia Impairment Scale (CIS):** (Bird *et al.*, 1993)

This is a 13-item global functioning questionnaire that can be administered by lay interviewers to assess (1) interpersonal relations, (2) broad psychopathological domains, (3) functioning in job or school work and (4) use of leisure time. It has a six month time frame and there are parent and child versions. It collects the information needed for making a judgment about functioning by asking a series of questions of the respondent. The scale has been tested in a combined clinic and community New York sample for internal consistency, as well as construct validity (Bird *et al.*, 1993) The scale has also been tested in a multi-ethnic community sample as part of the Methods for the Epidemiology of Child and Adolescent Mental Disorders (MECA) for test retest reliability and construct and concurrent validity (Bird *et al.*, 1996). As Table 1 shows, lower reliability and validity were reported for youth than for parents. Three of the items measure broad psychopathological domains rather than functioning, although the scale's psychometric properties were not affected by excluding these items. The main advantage of the scale is its ease of administration and brevity for both research and clinical settings. At least one item of the CIS can be construed as assessing participation restriction (*getting along with the family*), and two tap activity limitations (*getting involved with sports or hobbies; problems doing school work/job*).

## **2. Domain-Specific or Multidimensional Scales**

These are measures that rate a child's functioning in different domains and provide separate scores for each area of functioning. Children can behave differently in different contexts or situations, and it is possible for a child's functioning to be impaired, or restricted, in one area of functioning but not in another. Multidimensional scales are useful for planning both treatment and prevention, since they provide a basis for an action plan for different contexts; for example, for home and school. A serious disadvantage of many of these measures is that they mingle psychiatric symptoms with functioning items, not permitting the differentiation of these two important constructs.

### **2A. Child Behavior Checklist (CBCL) Social Competence Scale:** (Achenbach & Edelbrock, 1983; Achenbach, 1991a)

The CBCL is a questionnaire designed for parents about their child. One section deals with emotional and behavioral problems, and the other, called "Social Competence," with the child's functioning in three areas: school functioning, activities outside school (sports, hobbies, organizations, etc.), and social relationships with siblings, other children, and parents. Functioning

is judged relative to that of other children of the same age. The latest testing of the scale showed good test retest reliability (Achenbach, 1991a). Comparisons between 4,455 referred children with 2,368 non-referred obtained from a national probability sample in the United States showed that 21% to 39% of the variance in adaptive functioning scores was accounted for by referral status. This widely-used scale has several advantages. It has been tested in several ethnic groups and ethnicity has accounted for less than 1% of the variance in social competence scores. It has been translated into more than 50 languages and it has been tested and used all over the world. The Social Competence section takes 5-10 minutes to complete, and comparison with US national norms is possible. However, it has limitations:

(1) testing of the scale among Puerto Rican children showed that while a significantly larger portion of community children had scores in the clinical range compared to similar mainland US children (Achenbach *et al.*, 1990), this was mainly because of lack of involvement in activities by many of the children, resulting from a lack of resources and opportunities. This lack of opportunity translated into low scores on the Activities subscale. The CBCL assumes normal (by American standards) access to opportunities for schooling, social relationships, and activities, and may not reflect the true functioning of impoverished children with few opportunities to engage in hobbies, sports or organizations, or children living in highly dysfunctional families, or without access to adequate schooling.

(2) The CBCL is not designed to record extremes of deviation, and so may not adequately reflect the level of functioning of a very sick, disturbed or developmentally delayed child.

(3) The CBCL does not collect new information about a child; it assumes that the person completing it already knows the child well. Thus, its use by social workers, residential care workers, disability assessors, and other professionals is only as valid as their knowledge of the child.

Similar scales of adaptive functioning are included in the parallel versions of the CBCL that exist for self-report by children aged 11 and over; for young adults; and for parents of toddlers. The version for teachers discussed below also has many similarities, but it also collects some importantly different information.

## **2B. Teacher Report Form (TRF): (Achenbach, 1991b)**

This version of the CBCL is designed for teachers to complete about a child. It assesses the child's school functioning, focusing on both strengths and weaknesses. It is completed by the child's teacher or a counselor who knows the child in the school setting. The Social Competence section records information about the child's academic performance, scores on standardized tests, grade for age, special classes for educational or behavioral purposes, disabilities or handicaps, and good and bad points in the eyes of the teacher. As with the other versions of the CBCL, the instrument has several advantages. Test retest reliability of the scale is good, and comparison of 2,815 nonreferred with 1,391 referred children showed that 23% to 36% of the variance in adaptive functioning scores was accounted for by referral status. It has been tested in various ethnic groups, and ethnicity accounted for less than 1% of the variance in adaptive functioning scores. It provides parallel information to the CBCL, adding information that the parent may not know about the child. Its main limitation is that it requires the teacher to know the child well, which can be a problem early in the school year, or for high school students who have multiple teachers.

Ideally it should be completed by several school personnel, including the child's counselor.

## **2C. Social Adjustment Inventory for Children and Adolescents (SAICA):** (John *et al.*, 1987)

The SAICA is a semi-structured interview for school age children, or for parents about their children, for use by clinicians or trained interviewers. It covers four role areas: functioning at school, in spare-time activities, with peers, and with siblings and parents. It generates 12 subscales, a global score, and an interviewer's global assessment. Children with diagnoses have been found to have higher parent report SAICA scores than normal controls (John *et al.*, 1987). Other studies (Biederman *et al.*, 1991) have found that children who meet criteria for ADHD were more likely to be impaired than children with no diagnoses. In addition, children with comorbid psychiatric disorders have also been found to be more impaired than controls and more impaired than ADHD-only children.

Unlike the CBCL, the SAICA does not assume knowledge of the child, but provides a standardized way of collecting such information. Not surprisingly, it takes longer to complete (about 30 minutes). The main disadvantages of the scale are: (1) The child version does not discriminate children with psychopathology from controls. (2) The reliability of the instrument has not been reported. (3) It has not been tested with different ethnic groups. (4) There are substantial overlaps in its item content with what are usually regarded as being symptoms of disorders. The instrument, however, is unusual in paying attention to developmental differences in appropriate functioning; e.g., it contains an optional section on boyfriends and girlfriends for use with older children. In addition, the semi-structured interview format may fit into clinical settings better than paper and pencil measures.

## **2D. Vineland Adaptive Behavior Scales:** (Sparrow *et al.*, 1984)

The Vineland is a semi-structured interview measure of personal and social sufficiency for administration by a trained interviewer or clinician to a parent or caregiver. There are three versions: a 297 item version, an expanded version of 577 items and a screener of 45 items (see below). A classroom version for teachers is also available. The instrument assesses adaptive behavior of individuals from birth to adulthood in the areas of communication, daily living skills, socialization, and motor skills. The Vineland collects information about the child using semi-structured questioning, and observation of the child in context. It has good test-retest reliability (Cicchetti *et al.*, 1991a; Cicchetti *et al.*, 1991b) and there are national US norms available which provide age standardized scores, as well as national percentiles and stanines.

The instrument has several advantages: (1) It is by far the best researched and most widely used of measures of adaptive functioning; (2) Because it was developed specifically to assess children with developmental delays, mental retardation, and other severe disabilities, it usefully discriminates among children at the lower end of the distribution of functioning; (3) It is developmentally sensitive, providing different items for different age groups, (4) There are translations of the instrument into Spanish and several other languages, (5) A study of the instrument in 70 children with special needs found no difference in scores among European, Asian, and Afro-Caribbean children (Bickerton *et al.*, 1995). The limitations of the instrument are: (1) some of the items are culture bound (e.g., *table manners, orders own meal in a restaurant*), and (2) items in the maladaptive domain largely overlap with psychiatric symptoms.



## **2E. Vineland Screener:** (Cicchetti *et al.*, 1991a)

This is a 45 item version of the full Vineland Adaptive Behavior Scales, designed for rapid administration in research settings. As with the full scale, it is a semi-structured interview administered by a trained interviewer, but it has the advantage that lay interviewers can be trained to administer it. The screener covers adaptive behavior of individuals from birth to adulthood in the areas of communication, daily living skills, socialization and motor skills. There are separate scale items for ages 0-2, 3-5, 6-12 and 13-18 and age standardized scores can be generated, as well as national percentiles and stanines. The scale scores of the screener correlate highly with equivalent scale scores on the full Vineland. Test retest reliability of the scale in a small sample showed very high reliability. It has the disadvantages of the full scale and all its advantages, with the additional advantage of being considerably shorter. However, the authors do not support its use outside research settings (Sparrow, personal communication).

## **2F. Child and Adolescent Functional Assessment Scale (CAFAS):** (Hodges, 1990)

The CAFAS is a schedule used to rate a child's functioning in eight areas: role performance at home, school and community, thinking (rational thought process), behavior toward self (appropriateness of the child's daily behavior), moods/emotions, self-harmful behavior and substance use. Caregivers are also rated on two scales: ability to provide for child's basic needs, and family/social support. Several examples of functioning at different levels of impairment are given, and for each subscale, the rater endorses the item that best describes the youth. The scores assigned to the descriptors selected in this way are summed to give subscale and total scores. Ratings are based either on the information that the rater already has about the child, or on a psychiatric assessment, or on the *CAFAS Interview*, an instrument specially designed to collect the information needed to rate the CAFAS. Administration time for the latter is around 30 minutes, while completion of the CAFAS ratings takes around 10 minutes.

The CAFAS has been widely tested. Inter-rater and test-retest reliability in clinic samples is good (Hodges & Wong, 1996). The concurrent, predictive and discriminant validity of the instrument has been established in community samples (Hodges & Wong, 1997a). The instrument has been found to correlate moderately with the CBCL, the Burden of Care Questionnaire, the C-GAS, and the CAPA incapacity scales (see below), and to identify children with psychopathology according to the parent version of the Child Assessment Schedule. In addition, the CAFAS total score at intake has been found to predict service utilization at six and twelve months post intake (Hodges *et al.*, 1998). Higher impairment scores have been related to more restrictive use of services, higher costs, more bed days and more days of services (Hodges *et al.*, 1998). It also predicts other indicators of poor outcome, such as delinquency and absence from school (Hodges & Cheong-Seok, 2000).

The CAFAS has several advantages: (1) It has very good psychometric properties, being one of the few measures that has been tested longitudinally and found to have good predictive validity. (2) If the interviewer or clinician knows the child well it takes only about 10 minutes to rate. The limitations of the scale are: (1) Most of the scales, except for the role performance scale and the caregiver scales, intertwine psychiatric symptoms with functioning items, not permitting the differentiation of these constructs. (2) it has been tested only on Anglos and African Americans (for whom no separate psychometric data are available).

## **2G. Behavioral and Emotional Rating Scale (BERS):** (Epstein & Sharma, 1997)

This 52 item rating scale of a child's social and emotional strengths is designed for completion by an adult who knows the child well (Epstein & Sharma, 1997). It assesses 5 dimensions of childhood strengths: interpersonal strengths (e.g., ability to control emotions); involvement with family; intrapersonal strengths (child's outlook on his/her accomplishments and competence), school functioning, and affective strengths (e.g., asks for help, accepts a hug). It has a three month time frame. The scale has been tested in a combined clinic and community sample of Caucasian US children. Inter-rater and test-retest reliability coefficients were excellent and the scale discriminated well between children with and without disabilities. It is designed for use in special education and child welfare, as well as in mental health settings, to target goals for individual treatment plans. It is one of the few scales to focus on the child's strengths rather than weaknesses. Its main limitations are; 1) it has not been tested in diverse ethnic groups, and 2) some of the language in the scale requires fairly high reading and comprehension skills.

## **2H. Instrumental and Social Competence Scale:** (Beiser *et al.*, 1993)

This social adaptation scale is designed to measure a child's ability to get along with friends, and to perform appropriately at school. The ISCS takes the form of 25 items for parents or caregivers, and 21 items for children, which generate an instrumental scale that measures the child's competencies in school and work, and a social competence scale that measures the child's ability to get along with friends, peers and other people. The scale has been tested on US and Canadian samples of Anglos, African Americans, American Indians and Puerto Ricans, and the two subscales have been found to have high internal consistency in these groups (Beiser *et al.*, 1993; Goodman *et al.*, 1998). The instrument is short (it takes 5 minutes to administer) and has adequate convergent validity. With two exceptions, all correlations between the total score on the ISCS and teacher ratings and achievement tests for were significant in samples of native and non-native Canadian children in different grade levels. The main limitations of the scale are: (1) It measures only two domains, school and social relations, and thus provides no estimation as to how the child functions in other domains such as at home and his or her community or neighborhood (2) No test-retest data are available.

## **2I. Child Health and Illness Profile-Adolescent Edition (CHIP-AE)**(Starfield, 1994)

The CHIP-AE is a 219-item questionnaire describing current health status in six broad domains and 20 subdomains. The 6 broad domains are labeled Satisfaction (satisfaction with one's health, well-being, self-esteem), Discomfort (physical and psychological symptoms that interfere with well-being; the subdomains are labeled Physical Discomfort, Emotional Discomfort, Limitations of Activity), Resilience (Physical Activities, Problem-Solving, Home Safety and Health, Family Involvement), Risks (states and behaviors known to increase the likelihood of ill health or injury), Disorders (mental and physical illnesses, injuries, impairments), and Achievement (Academic Performance, Work Performance). The CHIP-AE is available in formats for parent report or self-report, and take 30-45 minutes to complete. It can be administered over the telephone, but this application has not been validated. Scoring can be done manually or using a computer-based program. Each subdomain has a mean of 20 and a standard deviation of 5. The CHIP-AE requires a fifth-grade reading level.

Scores on 4 of the domains (Satisfaction, Discomfort, Risk, Resilience) have been aggregated to provide 13 profile-types, ranging from Excellent Health (good score in all domains) to Poor on 3 or 4 Domains. The authors have tested the validity of these 13 profile-types against a range of constructs, including the other two domains (Disorders and Achievement).

Normative data, reliability, and construct validity on the CHIP-AE have been obtained from 4 public school samples from Baltimore, West Maryland, and Arkansas (total N=4,066) (Riley *et al.*, 1998). One-week test-retest reliability ranged across subdomains from ICC = .56 to ICC = .87. In a sample of 263 children interviewed using the DISC 2.3, 28% of those with a behavioral disorder and 20% of those with an emotional disorder were “poor” on 3 or 4 domains according to the CHIP-AE, compared with 9% of children with no psychiatric disorder (Riley *et al.*, 1998). Conversely, significantly fewer children with a psychiatric disorder than without scored in the “Excellent Health” range on the CHIP-AE.

Although there is extensive information on the CHIP-AE used with white and African American youth, very little is known about how it functions with Hispanic or other minority groups.

## **2.J. The Vanderbilt Functioning Indexes for Parent and Youth (VFI-P, VFI-Y) (Bickman *et al.*, 1998a)**

These two questionnaires, a 24-item one for parents and a 15-item one for youth, were developed in the context of a large services intervention study, the Fort Bragg Program Evaluation Project. Half of the study sample (total N=984) was randomly selected to derive the measures, and the other half was used to test the scales’ validity. Several measures of functioning used in that study, including the CBCL and its version for adolescents, the Youth Self Report (Achenbach & Edelbrock, 1991), were the sources of the items. Two different measures of functioning were used to measure criterion validity: teacher ratings of competence from the TRF, and provider reports of functioning on the GLOF. Two other validity criteria were also available from the Fort Bragg study: the total cost of treatment from intake to six-month follow-up, and whether the child had needed out-of-home placement at any time during that six months.

Twenty-four items were selected for the VFI-P, and the 15 items selected for the VFI-Y, using the criterion that each contributed significantly to the prediction of at least one of the four validity criteria. They were divided into 5 categories (defined *a priori*): antisocial behavior, problems at home, problems at school, problems with peers, and self-harm. Total score on the VFI-P and VFI-Y predicted the criterion measures better than the CBCL problem or competence scales, and as well as the much longer CAFAS.

There is no test-retest reliability information on the VFI. The sample on which it was developed and tested was predominantly White (71%), and as the children of military personnel were far from poor (average family income \$20,000-\$40,000). Also, they were all recruited when they attended a mental health facility. Thus, we do not yet know whether the high criterion validity reported (Bickman *et al.*, 1998a) is sustained in general population samples.

## **3. Measures of Symptom-Specific Impairment**

These measures determine the degree of disability associated with a specific symptom or diagnosis, and therefore need to be given in conjunction with a psychiatric assessment. Both the

Diagnostic Interview Schedule for Children (DISC) (Fisher *et al.*, unpub (1996)) and the Child and Adolescent Psychiatric Assessment (CAPA) (Angold *et al.*, 1995) include a series of questions that ascertain whether the symptoms reported to be present led to impairment in school functioning or relationships with family or peers. An advantage of these combined assessments of psychopathology and functional impairment is that, by distinguishing carefully between *distress associated with the symptom* and *impairment associated with the symptom*, they make it possible to determine the effect on prevalence estimates of adding or not adding impairment in functioning to the diagnostic criteria. Since the DSM-IV requires either distress or impairment, whereas some service agencies and entitlement programs require both, the ability to make this distinction is valuable. This feature is also useful for research, since the effect of impaired functioning on persistence of a disorder, or on time to recovery, needs to be established in future studies.

### **3A. Child and Adolescent Psychiatric Assessment (CAPA) Incapacity Ratings:** (Angold *et al.*, 1995)

These ratings identify areas of a child's life where functioning has been impaired because of his/her psychiatric symptoms, and the level of that impairment. When impairment in an area of functioning is reported, it is then attributed to one or more symptom areas. The term *incapacities* is used to distinguish these constructs from the World Health Organization's original ICIDH definitions of impairment, disability, and handicap. The areas covered by these interviewer-based rating are: withdrawal and or discord in relationships with parents, siblings, teachers and peers, and performance and behavior in home, school and elsewhere, including personal hygiene and self-care. Trouble with the law, treatment-seeking, and amount and type of treatment are also examined and linked to specific symptoms or diagnoses. This approach permits the respondent to attribute incapacity in a given area (e.g., school failure) to one or more symptom areas (e.g., to both anxiety and conduct disorder), and conversely to attribute incapacities in multiple areas to the same symptom (e.g., school failure and withdrawn relations with parents to depression). The CAPA incapacity ratings take little time to administer, and have good test-retest reliability (see Table 1 and (Angold & Costello, 1995)). No differences in reliability were observed between African American and White samples.

Validity has been examined in a large longitudinal community study, in which children have been assessed annually for four years (Costello *et al.*, 1996a). Contributing to the nomological net of validity criteria for the measure are the following: (1) Persistence: 26% of children with a diagnosis accompanied by functional impairment (i.e., SED) were SED a year later, compared with 4% of those who had only a diagnosis with no incapacity; (2) Family psychiatric history: SED children were 70% more likely to have a family history of mental illness than were children with only a diagnosis and 40% more likely than children with only incapacity (39% vs. 23% vs. 28%); (3) Service use: Children with SED were three times more likely to be in mental health care than children with either a diagnosis or incapacity alone (27% vs. 9% vs. 9%).

In this sample, African American children were significantly more likely to show incapacities than White or American Indian youth. This was most noticeable for school functioning, where they were particularly vulnerable to the disabling effects of disruptive behavior disorders. Boys in general were at higher risk for school incapacities, and girls for incapacities associated with family relations.

The instrument has not been tested in Hispanic samples, but work is in progress to translate it into Spanish, using an international bilingual team to ensure that it will be useful in different

Spanish-speaking countries throughout the world.

Use of the CAPA incapacity ratings is easiest in the context of the CAPA psychiatric interviews for parents and children (Angold *et al.*, 1995), which are designed to be administered by lay interviewers who have undergone a training course lasting two to four weeks. The additional training required to make the incapacity ratings is minimal.

### **3B. NIMH Diagnostic Interview Schedule for Children. Version IV (DISC-IV) Impairment Measures: (Fisher *et al.*, unpublished)**

Impairment questions are asked after each section of the DISC psychiatric interview of parent or child, if the child has a specified number of symptoms in a given diagnostic area. The DISC impairment measures assess whether the symptoms reported led to impairment in relations with family or peers, feeling upset, poor school functioning, treatment seeking, or trouble with the law (for externalizing symptoms only). The main advantages of the measure, besides the ones reported above, are its brevity and the fact that interviewers do not require lengthy additional training (as they would to administer, for example, the SAICA or the Vineland). A test-retest study of a prior version of the measure (DISC 2.3), which only covered impaired functioning with family, peers or in school, revealed that impairment ratings based on the parent interview were quite reliable, but those based on the child interview had satisfactory reliability only for the school setting. The psychometric properties of the impairment measures in the new version of the DISC, which has some additional items, have not yet been published. The main disadvantage of the measure is that the subject needs to make an attribution that the impairment is caused by the symptom or disorder under review. This requires a level of awareness about psychiatric symptomatology on the part of the parent or child that is not present in many cases. In fact, preliminary results of the DISC.2.3 version (Bird, personal communication) have shown that if parents or children report impairment for one set of specific symptoms they are more likely to report the same impairment for any other set of endorsed symptoms, suggesting that individuals can not easily differentiate impairment caused by one type of symptom versus another.

### **THE IMPACT OF CULTURE AND CONTEXT ON THE MEASUREMENT OF DISABILITY**

Functional adaptation is defined in terms of the way the person's role performance conforms to the expectations of his reference group (Horwitz *et al.*, 1996). Disability measures are thus based on an implicit standard: behaviors or roles which are normative to a given society or context (Katschnig, 1983). Given this contextual definition, one would expect definitions of "disability" to vary across different cultural and/or socioeconomic groups. Evidence from the 1985 Puerto Rican survey (Bird *et al.*, 1988) supports the view that assessments of social adaptation must be contextually developed. Even after matching children for age, sex and socioeconomic status, differences between Puerto Rican and Anglo children were observed, especially on the CBCL's Social Competence Scores (Achenbach *et al.*, 1990). Overall, social competence scores reported by adolescents and their parent and teachers were considerably lower for the Puerto Rican sample than for the Anglo sample. However, the Puerto Rican children reported more frequent contacts with friends and got along better with family and siblings. These last results may reflect the importance placed on close family ties and good interpersonal relations in Puerto Rican culture (Canino & Canino, 1982; Harwood *et al.*, 1995). It was reports of involvement in sports, hobbies,

organizations, or jobs that were significantly lower for Puerto Rico's children, contributing to their lower Social Competence Scores. The lack of resources in poor communities in Puerto Rico limit children's access to hobbies and formal sports programs. High unemployment rates on the island limit the availability of jobs. These social conditions affect younger age groups particularly, and probably account for these findings better than drawing the conclusion that Puerto Rican children are less socially adapted.

While Puerto Rican children scored considerably lower than comparable Anglo children on the Social Competence Scale of the CBCL, (that is, as less competent), the opposite was found using two other measures of impairment: the non-clinician C-GAS and the CIS. In the MECA study (Shaffer *et al.*, 1996) Puerto Rican children were rated by both parents and lay interviewers on both the C-GAS and the CIS as less impaired than the comparable youth at the three mainland sites (Bird *et al.*, 1996). Bird *et al.* (1996) in reporting the psychometric properties of both scales, noted site differences particularly for the youth lay interviewer C-GAS, where most associations with the outcome measures were found to be significantly weaker in the Puerto Rico site. Site-specific correlations between predictors of impairment and clinician validators for the lay C-GAS and the CIS were all lower for the Puerto Rico site than on the mainland. The proportion of DISC psychiatric diagnoses accompanied by impaired functioning (using the parent lay interviewer C-GAS and a cut point of < 70) were 16.0% for Georgia, 19.9% for New Haven, Connecticut, 12.8% for Westchester, New York and 4.1% for San Juan, Puerto Rico. Low prevalences were also found in the Great Smoky Mountains study (Costello *et al.*, 1996a; Costello *et al.*, 1996b), and the Caring for Children in the Community study (Costello *et al.*, in press), which examined children in rural North Carolina. In these studies the prevalence of severe emotional disturbance in children aged 9-13, defined as having a DSM-III-R diagnosis and scoring < 70 on the C-GAS, was 4.2% for GSMS and 10.0% for CCC.

Further work from the GSMS has also identified significant differences in patterns of disability between majority and minority study participants. Minority respondents were somewhat more likely to report the presence of some disability. This was accounted for by higher probabilities of having school and peer disabilities. Minority children and adolescents were not more likely to have family disabilities.

Any measure of disability or adaptation must be carefully evaluated in its cultural and social context to avoid bias and misinterpretation of research results from studies using standard methods developed in one socio-cultural context. Many measures of impairment or social adaptation seem to be affected by cultural and contextual factors. The C-GAS, and possibly other measures requiring the interviewer to make a judgment about the extent of functional impairment, may be influenced by the rater's experience, prior knowledge of the child, and expectations of what constitutes impairment and social adaptation in a particular culture. The CIS, and other measures that ask questions of the child, parent, or caregiver about the child's functioning are likely to be influenced by that individual's perceptions of what constitutes adaptive functioning. Furthermore, similar considerations must be borne in mind in relation to other demographic characteristics of study participants. For instance, in the GSMS, there was a 30% excess of any disability in boys, entirely accounted for by the presence of much higher rates of educational disabilities in the boys. In boys educational disabilities were the most common form of disability, while in girls family disabilities were most common.

### HOW MANY KINDS OF DISABILITY?

The scales reviewed here vary in the number of constructs that they set out to measure, from the single factor of the CIS to the three factors of the SAICA, to the CHIP-AE's ten factors. In many cases it is not clear how authors arrived at the number of constructs that they selected.

A factor analysis of the CAPA impairment data by Ezpeleta et al. (Ezpeleta *et al.*, submitted) provides support for both a one factor and a three factor solution. A single factor solution provided a reasonable description of the relationships among different areas of impairment, whereas all the multifactorial solutions were more or less unstable, despite relatively large samples. The most satisfactory three factor solutions broadly indicated that sub-domains of school disabilities, family disabilities and peer disabilities could be distinguished. The content of these three factors is quite similar to the three factors of the SAICA (labeled "task performance", "family relations", and "spare time sociability"). Solutions with higher numbers of factors were very unstable, and there was no support for the existence of as many as 10 factors. Nor was there any evidence for separate factors for socially withdrawn, as opposed to socially disruptive, disability. These two aspects of the way the CAPA assesses social disability consistently loaded together within the school, family and peer dimensions, and showed no sign of consistent separation even in models with larger numbers of factors.

### HOW WELL DO CURRENT MEASURES ASSESS ICDH-2 CONSTRUCTS?

In order to be useful for assessing disability in ICDH-2 terms, measures need to assess some or all of the constructs involved in functional impairment, activity limitation, and participation restriction. They need to do so in such a way that the different types of impairment can be distinguished, and attributed to different symptoms or disorders.

All but one of the measures described permit some sort of assessment of some aspects of ICDH-2 impairment (the exception is the C-GAS, which simply provides a global rating of impairment that combines all three types of functioning). Most contain at least one item in each of the three areas. For example, one of the shortest measures, the Columbia Impairment Scale, has items that tap functioning (*getting into trouble, having fun*), problems with activity limitations (*getting involved in activities like sports or hobbies, having a problem with doing your school work/doing your job*), and participation restriction (*getting along with family, friends*). However, the distinctions are often not clear; it is perfectly reasonable to argue that *problems getting along with family or friends* is an activity limitation, not a participation restriction. The reality is that the distinction is not an easy one to make, and future attempts to capture the different domains of impairment in an assessment instrument will need to be careful not to distort the underlying "reality" by measuring only what fits the framework and ignoring areas of functioning that do not fit it easily. The CAPA and DISC-IV both maintain a clear distinction between symptoms and disabilities, and probably come closest to the ICDH-2 conceptualization in this regard.

### HOW APPLICABLE ARE THESE MEASURES TO OTHER TYPES OF ILLNESS?

We have reviewed measures of disability in the context of psychiatric disorder because it is the context in which much of the work on concept formation and instrument development has taken place. But the inspiration for several measures came from earlier work to measure disability

associated with other types of illness (e.g., the Global Assessment Scale (Endicott *et al.*, 1976)) or developmental handicap (e.g., the Vineland).

The structure of ICDH-2's approach to disability creates no boundary between "mental" and "physical" illness. As noted earlier, most psychiatric disorders, and many "physical" ones, are still diagnosable only in terms of *impaired functions and structure*. As this changes, however, the basic structure of the ICDH-2's approach should continue to be applicable to both "mental" and "physical" illness.

#### **STRUCTURED INTERVIEW METHODS FOR COLLECTING INFORMATION ON CHILD AND ADOLESCENT DIAGNOSIS AND DISABILITY**

A basic distinction has arisen between two different strategies for structuring information coverage and defining ways to collect relevant information. These two methods have been dubbed "interviewer-based" (or sometimes "investigator-based") and "respondent-based" (Angold *et al.*, 1995). This distinction comes down to a difference in *what* is structured, or the level at which information is structured. In an interviewer-based interview, the mind of the interviewer is structured. In essence, the interview schedule serves as a tool to guide the interviewer in determining whether disability is present, but the interviewer makes the decisions, on the basis of information provided by the child or adult. In order to reduce idiosyncrasies in these interviewer judgments, definitions of items to be coded are provided, and the interviewer is expected to question until s/he can decide whether the phenomena described meet these criteria. Psychiatric interviews of this sort were the first to be developed, since they sprang naturally from clinical practice. Early examples of this type interview were the PSE (Wing, 1974) and the Reynard (Guze *et al.*, 1969) for adults, and the Isle of Wight interview for children (Graham & Rutter, 1968; Rutter & Graham, 1968). In each of these cases, *clinicians* were expected to conduct the interviews, since it was felt that only they had the necessary training and experience to be able to decide about the presence or absence of symptoms, even when quite detailed definitions were provided.

Although the PSE and Isle of Wight interviews were used extensively in moderately-sized epidemiological surveys, it was clear that the use of clinician interviewers created both logistic and budgetary problems. Very large scale epidemiological studies, such as the Epidemiologic Catchment Area (ECA) studies (Regier *et al.*, 1984) mandated the use of non-clinician ("lay") interviewers. However, it was felt that such interviewers would be incapable of making the judgments required by interviewer-based interviews, so, respondent-based psychiatric interviews were developed, following methodologies used, by political and marketing surveys. In a respondent-based interview, it is the *questions* put to the subject that are structured, and the interviewer makes no decisions about the presence of symptoms. Prescribed questions are asked verbatim in a preset order, and the interviewee's responses are recorded with a minimum of interpretation or clarification by the interviewer. Information variance due to variability in interviewing style or content is thus minimized. The Diagnostic Interview Schedule was the paradigmatic example in adult psychiatry (Robins *et al.*, 1979). The original Diagnostic Interview Schedule for Children and Adolescents (DICA) was the first child interview of this sort (Herjanic *et al.*, 1975; Herjanic & Campbell, 1977), although it has progressively been transformed into an interviewer-based interview. NIMH sponsored the development of the respondent-based Diagnostic Interview Schedule for Children (DISC), beginning in the early 1980s (Costello *et al.*,



1984; Edelbrock & Costello, 1990). The obvious difficulty with such an interview is that, although one knows exactly what has been asked in each interview, and exactly what was answered, there is no control over differences in how subjects interpret questions or respond to them.

It is important to be aware, however, that the goals of these two interviewing strategies are the same - to reduce information variance as much as possible. Strategies for data combination to produce diagnoses and scale scores have also shown considerable convergence, with computer-scoring emerging as a key method for both respondent-based and interviewer-based interviews. It is also important to bear in mind that all of the interviews considered here are moving targets. Interview developers and users are constantly modifying and updating them in response to changes in nosological systems, the requirements of particular studies, and increasing experience with the strengths and weaknesses of their own and other's measures.

It is perhaps worth noting here that respondent-based interviews have often been referred to as "highly structured", while interviewer-based interviews have been called "semi-structured". These are misnomers, since the issue is not *how much* structure is present, but *what* is structured -- the questions or the definition of the items.

#### **ASSESSING DISABILITY USING STRUCTURED INTERVIEWS**

The variability in how disability is assessed using structured interviews of children does not depend solely on whether the interview is respondent-based or interviewer-based. It is also associated with where in the interview process disability is assessed, and what aspects of it (impaired functioning, activity limitation, participation reduction). The main distinction is between coding the degree of disability associated with each reported symptom or symptom cluster (e.g, in the case of depression, having separate disability codings for anhedonia, sleep problems, guilt, etc.), versus coding a general level of disability associated with having any symptoms. A further distinction can be made between ratings made without reference to the symptoms to which the individual may attribute the disability, and those that seek an attribution. Examples of the first and second strategies can be found in the DISC, which collects disability ("impairment") ratings for each group of symptoms, and also a general rating. The CAPA is an example of an interview that collects general measures of disabilities ("incapacities") at the end of the whole psychiatric interview, but then for each incapacity asks the respondent to attribute it to one or more of 17 symptom areas.

An alternative strategy (frequently used, for example, with earlier versions of the DISC) is to collect symptom data and then to use an independent measure of disability.

## DISCUSSION DOCUMENT ON THE MEASUREMENT OF DISABILITY IN CHILDREN AND YOUTH

This Discussion Document has three parts. In the first part we discuss possible directions for further research needed to develop a measure of disability in children and youth. In the second part we discuss how such a measure might be used in the world of diagnosis and treatment, and in the third part we discuss the implications of the proposed revision of the role of disability in child and adolescent illness for further revisions of taxonomies such as the ICD and DSM.

### 1. FURTHER RESEARCH ON MEASURING DISABILITY IN CHILDREN AND YOUTH

A conclusion drawn from the review of existing measures of impairment or disability was that although many of them identify constructs similar to ICIDH-2's *functioning*, *activity limitations*, and *participation restriction*, they do not set out to structure their questions in these categories, and they do not cover all the relevant areas in a systematic fashion. For example, several measures contain an item referring to impaired school functioning, but it is often unclear whether one should be thinking about the child's limited ability to engage in expected activities at school (*activity limitation*), or about restrictions placed by others on the child's participation in normal school activities (*participation restriction*).

A review of measures in terms of the ICIDH-2 framework, however, begs the question of whether that framework actually fits children; whether it can be made to include the range of ways in which they can be disabled, and whether it includes aspects of disability that have no relevance for children and youth.

The first "research possibility" would therefore be to create a measure that made it possible to test the hypothesis that the ICIDH-2 framework fits children as well as adults. This would be done using the standard methods of instrument design: the first step, a review of relevant instruments, has been done; this would be followed by focus groups with experts, disabled children and caregivers, stakeholders in the use of such a measure, and so on. From these two steps a preliminary list of constructs would be generated, and turned into questions to be tested in relevant samples. Subject debriefing would provide additional feedback about the extent to which the instrument covered relevant constructs. The results of this process would lead to a revision of both the instrument *and* the taxonomy.

The process of testing an instrument in relation to a taxonomy is thus one of successive approximations (witness the successive editions of the ICD and DSM and the instruments developed for use with sections of each). The first measure will test the fit of the model, and in turn will suggest ways in which the model needs to be adapted.

The technology of instrument development and testing is quite advanced nowadays, and there are companies that will undertake the whole process (for a price). Alternatively, academics with an interest in the construct of disability or its consequences or prevention might well undertake the task of instrument development in the service of that interest. (Most of the standard psychiatric and psychological assessment tools were developed in this way; professional instrument development is now common in psychology, but still quite rare in psychiatry.)

Given that the goal is to have a framework for describing and measuring disability of all sorts,

not just that associated with psychopathology, it would be a good idea to test successive stages of the measure on a wide range of disabled populations. Engaging care providers and potential users of the measure in all the relevant areas (pediatrics, education, welfare, charitable organizations, etc.) will be a very large undertaking, especially as several versions of the measure will probably need to be tested.

A complication that is central to work involving children is the need to ensure that the measure is appropriate for children across a specific developmental range. This means that someone has to decide whether to aim for a single measure of disability that will be appropriate from birth to adulthood, or whether to have two or more measures tapping disability at different stages; for example, in infancy, pre-school, childhood, and adolescence. The medical and psychiatric taxonomies have gone for the first, “development-free” option to define the various disorders, but the instruments used to measure the constructs have mainly had restricted age-ranges. For example, most measures of psychiatric disorders for adults have a lower age limit around 17 (although the UM-CIDI is being extended down to age 13). Psychiatric instruments for child tend to be described as applicable to the age range 8 or 9 to 17 or 18 (although some, like the CAPA and the CBCL, have “young adult” versions). Some measures (CBCL, CAPA) also have downward extensions calibrated to the problems of preschool children. So far, the psychiatric problems of infants have been assessed using different methods, and, indeed, have their own (Zero to Three) taxonomy. However, there is a lot of developmental territory even between ages 9 and 17, and it is not yet clear that one instrument to measure disability can operate reliably and validly across the full range.

Once the preliminary stages have been completed and a first draft of a measure (or a set of stage-specific measures) developed, the next question is where to find research populations for the testing phase. In the United States, one possibility might be to invoke the involvement of one or more of the physician research networks such as ASPEN or PROS for primary care medicine, or (for psychiatry) PRN. When one of these networks adopts a research project, each physician will complete a small number of assessments and send the results to a central site to be tabulated and scored. This can produce a large and sociodemographically diverse sample at relatively low cost. Validity studies would need to use a range of clinical settings, and a control group, to test whether the measure actually identified children with known disabilities. Other countries will need to develop other methods for recruiting children with a wide range of illness and disability.

## **2. FUTURE DIRECTIONS**

### **A. Use of the disability construct in the assessment of need, utilization, outcomes, and costs.**

An extensive period of testing will be needed before any measure will be accepted for use in the “real world”. This is particularly true in the case of needs assessment, where the assessor’s judgment can mean, literally in some cases, life or death for a child. In other cases it can mean the right to income supplements, access to special educational services, or approval of treatment by a managed care company or insurer. At the level of government and public policy, the prevalence of “need for services” in a community can dictate the amount of financial help to which the community may be entitled from a range of governmental agencies.

After a decade of struggling to use diagnosis as the criterion of “need for services” its weakness for such a purpose has been amply demonstrated. An example is mental health, where, for example, an anxiety disorder diagnosis can carry with it any level of disability from mild distress to crippling functional impairment, activity limitation, and participatory restrictions. It is absolutely critical that an acceptable measure of disability be developed to be used alongside diagnostic assessments.

It is worth noting that need, utilization, outcomes, and costs are not only reasons for developing both a taxonomy and a measure of disability; they are also the criteria by which a measure’s validity can be evaluated (Bickman *et al.*, 1998a). Children who endorse the categories described in the taxonomy, and who have high scores on the measure that incorporates them, should have diagnoses (“service needs”), use services, improve with appropriate treatment, and have high medical costs. Furthermore, if the taxonomy and measure are to add anything to the standard diagnostic system, they need to predict need, use, outcome and costs *controlling for* the severity of the diagnosis. Finally, many treatments would be expected to reduce disability as well as symptoms and distress, so that disability may appropriately be included in the outcome measures for treatment trials. (This has an implication for instrument development; if the treatment is short it might be necessary to develop two parallel versions of the measure so that

## **B. Use in revisions of DSM and ICD**

One of the most useful roles that standardized instruments serve is to enforce greater clarity of conceptualization in the taxonomy. For example, the DSM-IV definition of Oppositional Defiant Disorder specifies “often loses temper”, “often argues with adults”, etc. How often is “often”? Researchers have had to search the developmental literature, and carry out epidemiologic studies, to establish norms and criteria for the level of “often” that is truly pathologic at different developmental stages and in different settings. Just as empirical studies using the DSM or ICD taxonomies have provided information that could be used to improve the next iteration of the diagnostic nosology, so the development of measures that operationalize the ICIDH-2 for children will help to sharpen the next version.

A second important use of a new measure of disability for children in revising the taxonomy of disease is that it will force researchers and clinicians to sharpen their skills in determining what the boundaries are between disorder and disability, thus improving the clarity of the next iterations. As discussed earlier, this is a slow process requiring constant iterations of conceptual revision and empirical research. Of course, improvements in the measurement of disability itself will also increase the ability to define and assess disease.

## **CONCLUSIONS ABOUT THE MEASUREMENT OF CHILDHOOD DISABILITY**

The assessment of functional impairment and disability as it applies to child mental illness is currently at the stage that the assessment of symptoms and diagnoses was twenty years ago: it has grown up haphazardly to meet the needs of different research and service delivery programs, in the absence of a generally-accepted conceptual framework. Now, twenty years later, we are well supplied with reasonably reliable measures of child psychopathology, ranging from 25-item screening questionnaires to detailed interviews, all using very similar (and to an increasing extent empirically-based) taxonomic principles, which have been sharpened by exposure to empirical

testing using a range of interviews and questionnaires. It is time for a similar development in the area of disability, and with the ICDH-2 framework much of the hardest conceptual work has already been done.

## Summary Table

Name of Measure	Key Reference	Requires knowledge of child? Y/N	Completed by... (Interviewer, parent, child, teacher)	Psychometric properties		
				Test-retest reliability Kappa or ICC	Discriminant validity	Concurrent/Criterion Validity
<b>1. Measures of Global Impairment</b>						
<b>Clinician C-GAS</b>	Bird et al, 1990	Y	Interviewer	ICC = .91	Significantly ( $p < .001$ ) discriminated between cases and non-cases using multi-variate definition of impairment 90.4% of cases were correctly classified	With Axis V, Pearson .82, with overall severity, .82, with Total CBCL, .65
<b>Lay, C-GAS</b>	Bird et al, 1996	Y	Interviewer	ICC = .75	Regression models showed C-GAS scores significantly ( $p < .001$ ) predicted service utilization	With Clinician's C-GAS, Pearson = .68, ICC = .65 for parent C-GAS
<b>CIS</b>	Bird et al, 1993	N	Parent or child	ICC = .89 for parent, .63 for child version	Children who had used MH services were significantly ( $p < .001$ ) more impaired than those who had not used	Comparison with clinician's C-GAS, Pearson = .63 for parent, .43 for child version
<b>2. Measures of Domain-specific Impairment</b>						
<b>CBCL</b>	Achenbach, 1991	N	Parent or youth	ICC = .range .83-.89		Criterion Validity: 21% to 39% of the variance in adaptive functioning scores accounted by referral status
<b>TRF</b>	Achenbach, 1991	N	Teacher	Pearson = range .78-.93		Criterion Validity: 23 % to 36% of the variance in adaptive functioning scores accounted for by referral status
<b>SAICA</b>	John, 1987	Y	Interviewer			Criterion Validity: Children with diagnoses had higher scores
<b>Vineland Scales</b>	Sparrow et al, 1984	Y	Interviewer	ICC = range .62 - .78 caregivers, .80-.90 parents		
<b>Vineland Screeners</b>	Cicchetti, 1991	Y	Interviewer	ICC = .98		Criterion Validity: Scale scores correlate between .89 to .98 with equivalent scale scores on full Vineland Scale
<b>CAFAS</b>	Hodges & Wong, 1996	Y	Interviewer	ICC = .77	High scores on CAFAS were significantly ( $p < .001$ ) related to several types of problematic behavior in children	First wave zero order correlations with Child Assessment Schedule = .59, with Burden of Care Questionnaire = .36, with CBCL = .42

### Summary Table

Name of Measure	Key Reference	Requires knowledge of child? Y/N	Completed by... (Interviewer, parent, child, teacher)	Psychometric properties		
				Test-retest reliability Kappa or ICC	Discriminant validity	Concurrent/Criterion Validity
<b>BERS</b>	Epstein & Sharma, 1997	Y	Parent, teacher	Pearson = .85-.99		Criterion Validity: children with emotional problems scored one standard deviation above the mean than normal controls.
<b>Instrumental Social Competence Scale</b>	Beiser et al, 1993	N	Parent, child		Convergent validity: With standard achievement tests and teacher assessment of performance; Zero order correlations ranged from .21 (teacher ratings of native children) to .65 (achievement tests and competence scores)	

### 3. Measures of Symptom-specific Impairment

<b>CAPA Incapacity Ratings</b>	Angold et al, 1995	N	Interviewer	ICC = .76		Criterion Validity: Children with DSM III R diagnosis were three times more likely to show incapacity as those without disorder
<b>DISC Impairment Measure</b>	Bird et al, in preparation	N	parent or child	K = range from .47 to .41, .34 for over anxious disorder, .39 for any anxiety		

## REFERENCES

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

Achenbach, T.M. (1991b). *Manual for the Teacher's Report Form and 1991 Profile*. University of Vermont Department of Psychiatry: Burlington, VT.

Achenbach, T.M., Bird, H.R., Canino, G., Phares, V., Gould, M.S., & Rubio-Stipec, M. (1990). Epidemiological comparisons of Puerto Rican and U.S. mainland Children: Parent, teacher, and self report. *Journal of the American Academy of Child and Adolescent Psychiatry* **29**, 84-93.

Achenbach, T.M. & Edelbrock, C. (1983). *Manual for the Child Behavior Checklist and Child Behavior Profile*. University of Vermont: Burlington, VT.

Angold, A. & Costello, E.J. (1995). A test-retest reliability study of child-reported psychiatric symptoms and diagnoses using the Child and Adolescent Psychiatric Assessment (CAPA-C). *Psychological Medicine* **25**, 755-762.

Angold, A., Messer, S.C., Stangl, D., Farmer, E.M.Z., Costello, E.J., & Burns, B.J. (1998). Perceived parental burden and service use for child and adolescent psychiatric disorders. *American Journal of Public Health* **88**, 75-80.

Angold, A., Prendergast, M., Cox, A., Harrington, R., Simonoff, E., & Rutter, M. (1995). The Child and Adolescent Psychiatric Assessment (CAPA). *Psychological Medicine* **25**, 739-753.

Beiser, M., Lancee, W., Gotowiec, A., Sack, W., & Redshirt, R. (1993). Measuring self-perceived role competence among first nations and non native children. *Canadian Journal of Psychiatry* **38,6 (August)**, 412-419.

Bickerton, W.L., Vostanis, P., Cumella, S., Chung, M.C., Doran, J., & Winchester, C. (1995). Adaptive functioning and behaviour of children with special needs: Comparison between ethnic groups. *Mental Handicap Research* **8**, 156-167.

Bickman, L., Lambert, E.W., Karver, M., & Andrade, A.R. (1998a). Two low-cost measures of child and adolescent functioning for services research. *Evaluation and Program Planning* **21**, 263-275.



Biederman, J., Newcorn, J., & Sprich, S. (1991). Comorbidity of attention deficit hyperactivity disorder with conduct, depressive, anxiety, and other disorders. *American Journal of Psychiatry* **148**, 564-577.

Bird, H.R., Andrews, H., Schwab-Stone, M., Goodman, S., Dulcan, M., Richters, J., Rubio-Stipec, M., Moore, R.E., Chiang, P.H., Hoven, C., Canino, G., Fisher, P., & Gould, M.S. (1996). Global measures of impairment for epidemiologic and clinical use with children and adolescents. *International Journal of Psychiatric Research* **6**, 295-307.

Bird, H.R., Canino, G., Rubio-Stipec, M., Gould, M.S., Ribera, J., Sesman, M., Woodbury, M., Huertas-Goldman, S., Pagan, A., Sanchez-Lacay, A., & Moscoso, M. (1988). Estimates of the prevalence of childhood maladjustment in a community survey in Puerto Rico: The use of combined measures. *Archives of General Psychiatry* **45**, 1120-1126.

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

Bird, H.R., Shaffer, D., Fisher, P.W., Gould, M.S., Staghezza, B., Chen, J.Y., & Hoven, C. (1993). The Columbia impairment scale (CIS): Pilot findings on a measure of global impairment for children and adolescents. *International Journal of Methods in Psychiatric Research* **3**, 167-176.

Bird, H.R., Yager, T.J., Staghezza, B., Gould, M.S., Canino, G., & Rubio-Stipec, M. (1990). Impairment in the epidemiological measurement of childhood psychopathology in the community. *Journal of the American Academy of Child and Adolescent Psychiatry* **29**, 796-803.

Canino, G. & Canino, I. (1982). Family therapy: A culturally syntonic approach for Hispanics. *Hospital and Community Psychiatry* **33**, 299-303.

Cicchetti, D.V., Sparrow, S.S., & Carter, A.S. (1991a). *Development and Validation of Two Vineland Adaptive Behavior Screening Instruments*. San Francisco, CA.

Cicchetti, D.V., Sparrow, S.S., Volkmar, F., Cohen, D., & Rourke, B.P. (1991b). Establishing the reliability and validity of neuropsychological disorders with low base rates: Some recommended guidelines. *Journal of Clinical and Experimental Neuropsychology* **13**, 328-338.

Costello, A.J., Edelbrock, C.S., Dulcan, M.K., Kalas, R., & Klaric, S.H. (1984). *Development and*

*Testing of the NIMH Diagnostic Interview Schedule for Children in a Clinic Population: Final Report (contract no. RFP-DB-81-0027).* NIMH Center for Epidemiologic Studies:Rockville, MD.

Costello, E.J., Angold, A., Burns, B.J., Erkanli, A., Stangl, D.K., & Tweed, D.L. (1996a). The Great Smoky Mountains Study of Youth: Functional impairment and Severe Emotional Disturbance. *Archives of General Psychiatry* **53**, 1137-1143.

Costello, E.J., Angold, A., Burns, B.J., Stangl, D.K., Tweed, D.L., Erkanli, A., & Worthman, C.M. (1996b). The Great Smoky Mountains Study of Youth: Goals, designs, methods, and the prevalence of DSM-III-R disorders. *Archives of General Psychiatry* **53**, 1129-1136.

Costello, E.J., Angold, A., & Keeler, G.P. (1999). Adolescent outcomes of childhood disorders: The consequences of severity and impairment. *Journal of the American Academy of Child and Adolescent Psychiatry* **38**, 121-128.

Costello, E.J., Angold, A., Messer, S.C., & Farmer, E.M.Z. (1996c). *Symptoms and impairment: Serious Emotional Disturbance in the Great Smoky Mountains study.* The Research and Training Center for Children's Mental Health:Tampa, FL.

Costello, E.J., Keeler, G.P., & Angold, A. (in press). Poverty, race and psychiatric disorder: A study of rural children. *American Journal of Public Health*

Edelbrock, C. & Costello, A.J. (1990). Structured psychiatric interviews for children and adolescents. In *Handbook of Psychological Assessment* (ed. G. Goldstein and M. Hersen), pp. 276-290. Pergammon Press:New York, NY.

Endicott, J., Spitzer, R.L., Fleiss, J.L., & Cohen, J. (1976). The Global Assessment Scale (GAS). *Archives of General Psychiatry* **33**, 766-771.

Epstein, M.H. & Sharma, J. (1997). *Behavioral and Emotional Rating Scale (BERS): A strength-based approach to assessment.* PRO-ED:Austin, TX.

Ezpeleta, L., Erkanli, A., Costello, E.J., Keeler, G., & Angold, A. (submitted). The epidemiology of psychiatric disability in childhood and adolescence.

Ezpeleta, L., Granero, R., de la Osa, N., & Guillamon, N. (In press). Predictors of functional impairment in children and adolescents. *Journal of Child Psychology and Psychiatry*

Fisher, P., Lucas, C., & Shaffer, D. (unpublished). *NIMH Diagnostic Interview Schedule for Children, Version IV*. 722 W. 168th Stret, New York, NY 10032.

Goodman, S.H., Hoven, C.W., Narrow, W.E., Cohen, P., Fielding, B., Alegria, M., Leaf, P.J., Kandel, D., Horwitz, S.M., Bravo, M., Moore, R., & Dulcan, M.K. (1998). Measurement of risk for mental disorders and competence in a psychiatric epidemiologic community survey: the NIMH Methods for the Epidemiology of Child and Adolescent Mental Disorders (MECA) Study. *Social Psychiatry and Psychiatric Epidemiology* **33**, 162-173.

Graham, P. & Rutter, M. (1968). The reliability and validity of the psychiatric assessment of the child: II. Interview with the parent. *British Journal of Psychiatry* **114**, 581-592.

Green, B., Shirk, S., Hanze, D., & Wasntrath, J. (1994). The children's global assessment scale in clinical practice: An empirical evaluation. *Journal of the American Academy of Child and Adolescent Psychiatry* **33**, 1158-1164.

Guze, S.B., Goodwin, D.W., & Crane, J.B. (1969). Criminality and psychiatric disorders. *Archives of General Psychiatry* **20**, 583-591.

Harwood, R.L., Miller, J.G., & Lucca Irizarry, N. (1995). *Culture and Attachment: Perceptions of the child in context*. Guilford:New York.

Herjanic, B. & Campbell, W. (1977). Differentiating psychiatrically disturbed children on the basis of a structured interview. *Journal of Abnormal Child Psychology* **5**, 127-134.

Herjanic, B., Herjanic, M., Brown, F., & Wheatt, T. (1975). Are children reliable reporters? *Journal of Abnormal Child Psychology* **3**, 41-48.

Hodges, K., Bickman, L., Kurtz, S., & Reiter, M. (1992). A multidimensional measure of the level of functioning for children and adolescents. In *A System of Care for Children's Mental Health* (ed. A. Algarin and R.M. Friedman), pp. University of South Florida:Tampa, FL.

Hodges, K. & Cheong-Seok, K. (2000). Psychometric study of the child and adolescent functional assessment scale: Prediction of contact with the law and poor school attendance. *Journal of Abnormal Child Psychology* **28**, 287-297.

Hodges, K., Doucette-Gates, A., & Liao, Q. (1999). The relationship between the Child and Adolescent Functional Assessment Scale (CAFAS) and indicators of functioning. *Journal of Child and Family Studies* **8**, 109-122.

Hodges, K. & Wong, M. (1997a). Use of the Child and Adolescent Functional Assessment Scale to predict service utilization and cost. *Journal of Child and Family Studies* **5**, 445-467.

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

Hodges, K. & Wong, M.M. (1997b). Use of the child and adolescent functional assessment scale to predict service utilization and cost. *Journal of Mental Health Administration* **24**, 278-290.

Hodges, K., Wong, M.M., & Latessa, M. (1998). Use of the child and adolescent functional assessment scale (CAFAS) as an outcome measure in clinical settings. *Journal of Behavioral Health Services and Research* **25**, 325-336.

Horwitz, S.M., Hoagwood, K., Sparrow, S.S., & Triche, E. (1996). *Adaptive functioning: Its origin, meaning, and use in defining mental health problems, February 26-28*. The Research & Training Center for Children's Mental Health: Tampa, FL.

John, K., Davis, G.D., Prusoff, B.A., & Warner, V. (1987). The Social Adjustment Inventory for Children and Adolescents (SAICA): Testing of a new semistructured interview. *Journal of the American Academy of Child and Adolescent Psychiatry* **26**, 898-911.

Katschnig, H. (1983). Methods for measuring social adjustment. In *Methodology in Evaluation of Psychiatric Treatment* (ed. T. Helgason), pp. 205-218. Cambridge University Press: New York.

Lavigne, J.V., Binns, H.J., Arend, R., Rosenbaum, D., Christoffel, K.K., Hayford, J.R., & Gibbons, R.D. (1998). Psychopathology and health care use among preschool children: A retrospective analysis. *Journal of the American Academy of Child and Adolescent Psychiatry* **37**, 262-270.

Lyons, J.S., O'Mahoney, M., Miller, S.I., Neme, J., Kabat, J., & Miller, F. (1997). Predicting readmission to the psychiatric hospital in a managed care environment: Implications for quality indicators. *American Journal of Psychiatry* **154**, 337-340.

McArdle, P. & Gillett, T. (1997). Psychiatric in-patient treatment for children and adolescents in the UK: Criteria for admission. *Israel Journal of Psychiatry and Related Sciences* **34**, 195-199.

Nagel, E. (1957). Determinism and development. In *The Concept of Development* (ed. D.B. Harris), pp. 15-26. University of Minnesota Press:Minneapolis, MN.

Regier, D.A., Myers, J.K., Kramer, M., Robins, L.N., Blazer, D.G., Hough, R.L., Eaton, W.W., & Locke, B.Z. (1984). The NIMH Epidemiological Catchment Area Program: Historical context, major objectives, and study population characteristics. *Archives of General Psychiatry* **41**, 934-941.

Reich, W., Herjanic, B., Welner, Z., & Gandhi, P.R. (1982). Development of a structured psychiatric interview for children: Agreement on diagnosis comparing child and parent interviews. *Journal of Abnormal Child Psychology* **10**, 325-336.

Riley, A.W., Forrest, C.B., Starfield, B., Green, B., Kang, M., & Ensminger, M. (1998). Reliability and validity of the adolescent health profile-types. *Medical Care* **36**, 1237-1248.

Robins, L.N., Helzer, J., Croughan, J., Williams, J.B.W., & Spitzer, R.L. (1979). *The NIMH Diagnostic Interview Schedule (DIS): Version II*. National Institute of Mental Health:

Rutter, M. & Graham, P. (1968). The reliability and validity of the psychiatric assessment of the child: I. Interview with the child. *British Journal of Psychiatry* **114**, 563-579.

Setterberg, S., Bird, H., & Gould, M. (1992). *Parent and Interviewer Versions of the Children's Global Assessment Scale*. Columbia University:New York.

Settertobulte, W. & P., K. (1997). Gender-specific factors in the utilization of medical services during adolescence. *Journal of Adolescence* **20**, 121-132.

Shaffer, D., Fisher, P., Dulcan, M.K., & Davies, M. (1996). The NIMH Diagnostic Interview Schedule for Children Version 2.3 (DISC 2.3): Description, acceptability, prevalence rates, and performance in the MECA study. *Journal of the American Academy of Child and Adolescent Psychiatry* **35**, 865-877.

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

Sparrow, S., Balla, D., & Cicchetti, D. (1984). *Vineland Adaptive Behavior Scales: Interview Edition Expanded Form Manual*. American Guidance Service:Circle Pines, MN.

Spitzer, R.L. & Wakefield, J. (1999). DSM-IV diagnostic criterion for clinical significance: Does it

help solve the false positive problems? *American Journal of Psychiatry* **156**, 1856-1864.

Stanger, C. & Lewis, M. (1993). Agreement among parents, teachers, and children on internalizing and externalizing behavior problems. *Journal of Clinical Child Psychology* **22**, 107-115.

Starfield, B. (1994). The reliability and validity of a health status measure. *Medical Care*

Verhulst, F.C., Mart, L., Eussen, J.M., Guy, F.M., Berden, G., Sanders-Woudstra, J., & Van Der Ende, J. (1993). Pathways of problem behaviors from childhood to adolescence. *Journal of the American Academy of Child and Adolescent Psychiatry* **32**, 388-396.

Wakefield, J.C. (1997). When is development disordered? Developmental psychopathology and the harmful dysfunction analysis of mental disorders. *Development and Psychopathology* **9**, 269-290.

Wing, J.K. (1974). *Measurement and Classification of Psychiatric Symptoms*. Oxford University Press:Oxford.