The Developmental Behavior Checklist: The Development and Validation of an Instrument to Assess Behavioral and Emotional Disturbance in Children and Adolescents with Mental Retardation[1]

Stewart L. Einfeld[2]

School of Psychiatry, University of New South Wales

Bruce J. Tonge

Monash University Centre for Developmental Psychiatry

Describes the development and validation of the Developmental Behavior Checklist (DBC), a standardized instrument completed by lay informants to assess behavioral and emotional disturbance in children and adolescents with mental retardation (MR). Items describing common behavioral and emotional problems in this population were generated by extracting descriptions from 664 case files of children and adolescents with behavior disorders seen at a specialist developmental assessment service over 12 years. These items were reduced to a set of 96 items administered to a sample of 1,093 children and adolescents with mental retardation and then submitted to a principal

components analysis. Six interpretable and partly validated subscales were obtained which explained 36% of the total variance and had satisfactory internal consistency. Interrater and test–retest agreement were satisfactory for both total scale score and for scores on each of the subscales. Good evidence of concurrent validity was provided by substantial positive correlations between total scores on the DBC completed by lay informants and the ratings of experienced psychiatrists based upon interviews and scores on two standardized instruments that must be completed by health professionals. The discriminative validity of the total score as assessed by area under the ROC curve was excellent (92%). Standardized norms for the DBC are derived from an epidemiological study of behavior problems in children and adolescents with mental retardation undertaken in two Australian States. Norms are available for the mild, moderate, severe, and profound MR groups and for the MR population as a whole.

INTRODUCTION

Emotional and behavioral problems are a major source of additional handicap for children, adolescents, and adults with mental retardation (MR). They add to the suffering of the affected individual, cause distress to parents (Quine & Pahl, 1989), increase the likelihood of institutionalization (Ballinger & Reid, 1977), and reduce social integration and employment (Bruininks, Hill, Morreau, 1988).

Although it is established that intellectually disabled children have a higher rate of emotional and behavioral disturbance than other children (Corbett, 1979; Rutter, Tizard, & Whitmore, 1970), the relationship between mental retardation and behavioral or emotional problems is uncertain, as is the best approach to the assessment and classification of these problems (Jacobson, 1990; Scott, Swales, & Danhour, 1990). There is also a serious shortage of professionals with specialized training in the diagnosis and management of behavioral and emotional problems among persons with MR (Parmenter, 1988; Zarfas, 1988).

A standardized instrument for assessing behavioral and emotional disorders in children with MR would advance our knowledge of the relationship between behavioral and emotional disorders and MR and improve the clinical management of these problems. Such an instrument could be used to study the prevalence of behavior problems among children and adolescents with MR in the way that the Rutter scales (Rutter et al., 1970) and the Child Behavior Checklist (CBL; Achenbach & Edelbrock, 1983) have been used (Offord et al., 1987). It could identify associations between behavior problems and biological, psychological, and social variables, be used to investigate the etiology of behavior disorders and their response to treatment, and to further study and refine diagnostic classification systems.

A standardized assessment instrument would improve the clinical management of children with emotional and behavioral disorders and MR. Its use as a screening test could improve the efficient allocation of scarce mental health professionals by identifying individuals who required professional evaluation. If reliably completed by a lay informant, it could save valuable clinician time by providing a systematic and reliable summary of a wide range of symptoms. If administered routinely to those attending developmental pediatric clinics, it would provide a valuable history of behavior which would be of use in monitoring progress and treatment response.

This paper describes the development of such an instrument, the Developmental Behavior Checklist (DBC). It describes the way in which items of the DBC were generated and refined; the way in which their structure was investigated by a principal components analysis; a series of studies of the reliability and construct and criterion validity of the instrument, and various clinical and research applications.

A number of instruments for the evaluation of psychopathology in this population were already in existence when development of the DBC commenced in 1988. However, the authors believed that the limitations of these in meeting the goals described above justified the development of a new measure, a view shared by Aman and Singh (1988), Fraser and Rao (1991), and Sturmey, Reed, and Corbett (1991). Rather than provide a review of these preexisting instruments here, the reader is referred to the comprehensive review by Aman (1991a, 1991b). He identified 55 instruments for assessing emotional and behavior problems in both adults and children with developmental disabilities. He found (i) there was a lack of data evaluating the instruments' specificity and sensitivity in discriminating cases and noncases; (ii) few of the existing instruments had adequate standardization data in terms of different population groups with whom it might be used; (iii) there was a lack of instruments for assessing children relative to adolescents and adults; and (iv) a relative lack of recommended instruments for people with severe and profound MR.

Several widely used instruments were found by Aman to have significant limitations. The AAMD Adaptive Behavior Scales (Lambert & Windmiller, 1981; Nihira, Foster, Shellhaus, Lelond, 1975) lacked internal consistency data and taxonomic validity. The Vineland Adaptive Behavior Scales (Maladaptive Behavior Domain) (Sparrow, Balla, & Cichetti, 1984) were also regarded as lacking factorial/taxonomic validity. The Devereux Child Behavior Rating Scale (Spivack & Spotts, 1966) did not provide internal consistency and congruent validity data, the standardization sample was small, and the "intricate factor solution may prove to be unstable."

The Aberrant Behavior Checklist (ABC, Aman & Sing, 1987) is one instrument that appears to have similarities to the DBC in structure and properties. Although originally validated on adolescents and adults, it has been investigated for use in children. Rojahn and Helsel (1991) found that in children, the ABC retained its factor structure and criterion validity, but regarded its interrater reliability to be relatively low. No data on specificity and sensitivity were available. Although the ABC may prove useful in children as well as adults, the value of having different instruments of acceptable reliability and validity available, especially for research purposes, justified the development of the DBC (M. Aman, personal communication). The 96 items of the DBC may provide a more detailed description of psychopathology than the 58 of the ABC.

Currently available instruments may be evaluated in terms of a number of properties considered desirable in a new instrument.

The Rater. Some instruments require a professional to administer them, although information to assist the professional may be obtained from a lay informant. An example is the Vineland Adaptive Behavior Scale (Sparrow et al., 1984). The Schedule of Handicaps, Behavior and Skills (HBS-Revised; Wing, 1982) requires the rater to be a professional who has received special training in the use of the instrument. An instrument that is valid when completed by lay informants has obvious logistic advantages in large-scale surveys.

The Range of Psychopathology Covered. Some instruments are not intended to assess a wide range of psychopathology but rather specific disorders such as autism. One example is the Autism Behavior Checklist (Krug, Arick, & Almoud, 1980). The Diagnosis of Psychotic Behavior in Childhood (DIPBEC; Haracopis & Kelstrup, 1978) is limited to evaluation of "psychotic" behavior.

The Level of Intellectual Handicap for Which the Instrument is Appropriate. The Zung Self-Rating Scale (Adapted) (Lindsay & Mitchie, 1988) is suitable for the mildly but not the more severely retarded. Instruments for assessment of psychopathology in children and adolescents who do not have MR may also be considered for their applicability to this population. The CBCL (Achenbach & Edelbrock, 1983) is in wide use in investigating psychopathology in children and adolescents of normal intellect, and has been used in several studies of mildly intellectually handicapped children. It is not appropriate for the severely intellectually handicapped population (T. M. Achenbach, personal communication). This is because some of the items of the Child Behavior Checklist address cognitions that are too sophisticated for the under 50 IQ group. Many of the behaviors clinically observed in those with severe MR are absent from the Child Behavior Checklist. The Youth Self-Report is unsuitable for those of IQ below 50 as it requires

reading and writing skills, as well as an understanding of relatively sophis-ticated concepts. Another widely used instrument for the assessment of psychopathology in general child psychopathology, the Rutter scales (Rut-ter, 1967) do not include items seen more typically in disturbed intellectually handicapped children such as autistic symptoms.

The Age Group for Which the Instrument is Intended and on Which It is Standardized. For example, the Aberrant Behavior Checklist (Aman & Singh, 1987) was developed for use with adolescents and adults.

The Detail with Which Behavioral and Emotional Disturbance is De-scribed. Some instruments include items that are quite nonspecific. For example, the Scales of Independent Behavior (Bruininks, Woodcock, Weatherman, & Hill, 1984) has items such as "socially offensive behavior," which could include a wide range of different behaviors.

Our aim was to develop a standardized instrument that had these desirable properties. Its items would describe the child's disturbed emotions and behaviors specifically, and discriminate between behaviors attributable to developmental delay and behaviors which are evidence of emotional/be-havioral disorder. It would cover both emotional problems such as anxiety and behavioral items such as various aggressive behaviors.

It would be applicable to the moderately and severely intellectually handicapped, a group which clinical experience and the literature suggest has psychopathology that differs most from that seen in persons of av-erage intelligence (Russell, 1985), and a group that is least well served by existing instruments. Ideally, it would also be of use with mildly handi-capped children so that behavioral and emotional disorders could be assessed in children with the full range of mental retardation with which agencies work.

What should such an instrument asses? Should it evaluate symptoms, syndromes, or diagnoses? Should it measure categories or dimensions? If a categorical/diagnostic approach were used, would these be DSM, ICD, or other diagnoses? These issues are considered more fully by Einfeld and Aman (1995) who identified widespread doubts about the applicability of the DSM to the more severely intellectually handicapped (Aman, 1991a; Sovner, 1986). Factors giving rise to these doubts include a lack of data on the reliability and validity of DSM diagnoses in this population. One of the few studies to date (Einfeld & Tonge, 1991) reported low rates of interrater agreement for four major DSM-III-R diagnostic classes. An ex-ception is the group of Pervasive Development Disorders which achieved a level of overall agreement of 0.66 (kappa) in the DSM-III field trials (American Psychiatric Association, 1980). Studies of the prevalence of DSM-III diagnoses in apparently similar MR populations have produced widely discrepant findings e.g., Menolascino, 1988; Myers, 1987). This may

reflect a situation similar to that pertaining to the diagnosis of schizophrenia prior to the United Kingdom–United States Schizophrenia study (Karno & Norquist, 1989). Further, DSM diagnostic categories may not have practical utility. For example, Atypical Organic Brain Syndrome, the most common diagnosis of all in one survey (Rojahn & Helsel, 1991), has no diagnostic criteria, nor any established external validity. Sturmey (1993) has also noted that studies applying DSM criteria have all modified the criteria without evaluating the effects of these changes on their psychometric properties.

Rather than attempt to design an instrument to make DSM or ICD diagnoses, a different approach was adopted. We sought to describe the individual symptoms, list these, and then measure the properties of this symptom list. This may be called a *descriptive-empirical* approach. The advantages of such an approach were considered to be that it made no a priori assumptions about the nature or taxonomy of psychopathology in children and adolescents with MR. Further, other instruments based on the symptom checklist approach, such as the Rutter scales (1967) and the CBCL (Achenbach & Eldebrock, 1983) have been useful in a range of applications in general child psychopathology, the latter having been applied in more than 700 published studies (Achenbach & Brown, 1991). The descriptive-empirical approach also allows for the possibility that subgroupings of certain items may be found to correlate with DSM diagnoses. For example, Hepperlin, Stewart, and Rey (1990) extracted scores from the CBCL to diagnose depression. Ultimately, both the diagnostic (categorical) and the descriptive-empirical (dimensional) approach is necessary to advance knowledge and practice. The application of the DSM system was considered premature rather than inappropriate and comparison studies of the use of the DBC in relation to standardized diagnostic classifications are in progress.

METHOD

Item Development

The first step in item development was to identify behaviors and emotions of children and adolescents with MR which are of particular concern to parents, others, carers, teachers, and clinicians. This was done by surveying the clinic files of the Grosvenor Assessment Clinic, Sydney, Australia. This clinic has provided comprehensive developmental pediatric and psychological assessments of over 7,000 children and adolescents with mild, moderate, severe, and profound MR.

A comprehensive assessment protocol was completed for each child attending the clinic but no standardized behavioral instrument was used. Data on behavioral and emotional disturbances were recorded in reports from parents and teachers, and clinical observations of psychologists, developmental pediatricians, and psychiatrists. Of the most recent 4,500 files, 22% (n = 990) were coded according to Heber's classification (1961) as having a Behavior Disorder. When files containing inadequate descriptions of behavior were excluded, there were 664 files that contained detailed descriptions of behavioral problems. These descriptions were extracted, reduced to 135 descriptions of problem behaviors after eliminating synonyms. Two child psychiatrists, who are experienced in the psychiatric assessment of children with MR (B.T. and S.E.), determined which of these descriptions were symptoms of behavioral and emotional disorder according to a definition modified from Reid *et al.* (1978), itself adapted from Rutter *et al.* (1970), as follows:

> Where behavior and emotions are abnormal by virtue of their qualitative or quantitative deviance, and cannot be explained on the basis of developmental delay alone and cause significant distress to the child carers or the community, as well as significant added impairment, then these behaviors and emotions are regarded as disordered. Where the overall clinical presentation of the person shows evidence of such disturbed behaviors and emotions then the person is regarded as psychiatrically disturbed.

Behaviors that might have been readily attributed to developmental delay were also excluded or modified. For example, an item "soils outside toilet" was changed to "soils outside toilet, though toilet trained" to distinguish the child who has not yet achieved bowel control from the child with encopresis. Behaviors attributable to physical disabilities (e.g., "dribbles") were also excluded. A few other symptoms of clinical interest, namely, delusions, thought disorder, and abuse of drugs and alcohol, were added. This process produced 105 behavioral descriptions which were further refined by eliminating those with low interrater agreement in 200 files independent rated by two psychologists experienced in the assessment of children with MR. This resulted in 96 descriptions of behavior on which the average interrater .68 (by Cohen's kappa) and the test–retest reliability was .72 (also by kappa).

These 96 descriptions were translated into language suitable for parents, residential care workers, schoolteachers, or any person who was familiar with the intellectually handicapped child or adolescent. The extent to which the items were comprehensive to lay raters was estimated by a number of indices of readability (the Flesch Index 76.2; Flesch-Kincaid 4.8; Gunning-Fog Index 6.5; and the Flesch Grade Level 7.3), all of which indicated that the DBC could be understood by respondents who had completed a primary school education (Pfaffenberger, 1992). The authors reviewed checklists with informant parents to determine whether items were understood as they were intended, resulting in further refinements of item content.

The format of the items was a checklist using a structure adapted, with permission, from that of the CBCL (Achenbach & Edelbrock, 1983). Each item was scored on a 3-point scale: *Not true as far as you know* (0), *sometimes or somewhat true* (1), *often true or very true* (2). Items were listed in alphabetical order of their first letter. Items in the Primary Carer version were scored if present during the previous 6 months. The Teacher version contained 93 of the items of Primary Carer Version (items on sleep disturbance were deleted) and it was related to the past 2 months.

Factorial Structure of the Scale

The factorial structure of the scale was determined by a principal components analysis. The subjects for these studies were 1,093 children and adolescents with MR. These included subjects who were part of the epidemiological prevalence study and the reliability and validity studies described below (Einfeld, Tonge, & Florio, 1994). The group is broadly representative of the population for which the DBC would be used (see Tables I and II for age, sex and IQ distribution). The informants who completed the scale were parents, residential care workers, and nurses.

The principal components analysis extracted factors with an eigenvalue greater than 1, but only the first 6 factors were retained on the basis of a scree plot and their interpretability. These factors accounted for 32.6% of the total item variance. Although a seemingly modest proportion of total variance, this is what one would expect with a 96-item questionnaire that assesses a broad spectrum of behavioral disturbance (O'Grady, 1982). Summarized descriptions of the items loading greater than 0.35 on each factor after varimax rotation are shown below in Table III. The percentage of variance explained by each factor and its internal consistency as measured by Cronbach's alpha is shown in Table IV.

On the basis of the principal components analysis six scores were calculated by adding the scores for each item in each of the subscales. The items were not weighted by their factor loading to simplify scoring and because weighted and unweighted scores are highly correlated (Achenbach & Edelbrock, 1983; Streiner & Norman, 1989). A Total Behavior Problem Score (TBPS) was also calculated by adding all positive items. Although the analysis identified several independent dimensions the total score still provided an indicator of the severity of behavioral and emotional disorder. This is also the practice in scoring the widely used CBCL (Achenbach & Edelbrock, 1983).

Table I. Age and Sex of Subjects Used for Construct Validity Studies

Subjects	Age (years)			Total
	< 9	9–12	13+	
Male	165	141	349	655
Female	112	96	230	438
Total	277	237	579	1093

Table II. IQ Range of Subjects Used for Construct Validity Studies

IQ range	N	%
Mild retardation	271	24.8
Moderate retardation	350	32.0
Severe retardation	343	31.3
Profound retardation	33	3.0
Unknown (though < 70)	96	8.8

The subscales represented clinically relevant dimensions of behavior among children and adolescents with MR. Clinicians are frequently confronted with patients whose adaptation is impaired by the behaviors in the Disruptive subscale. One hundred sixty-six teachers completing the DBC-Teacher Version reported the level of disruption to the child's educational process on account of behavior problems on a 7-point scale comprising 3 subscales. The subscales measured the proportion of productive learning time lost, the increased share of time required for the child in the classroom, and the increased share of time required in non-"face-to-face" activity. There was a significant correlation between the level of disruption to the educational process as measured by the 7-point scale and the Disruptive subscale of the DBC ($r = .50, p < .0001$).

The behaviors of the "self-absorbed" subscale are more often seen in children and adolescents with severe MR (Einfeld, 1992), especially in institutions (Reid, 1983). The Communication Disturbance grouping of items is seen in verbal children with autism and is of increased severity in children with Williams syndrome, compared with IQ age- and sex-matched controls, $F = 7.0, p < .003$ (Einfeld, 1993). The Anxiety phenomena are also more prevalent in Williams syndrome than in the same controls, $F = 20.1, p \leq .001$.

Table III. Item Loadings on the Subscales of the Developmental Behavior Checklist[a]

Item	Loading	Item	Loading
Disruptive subscale		Self-absorbed subscale	
Manipulates	.661	Eats nonfood	.603
Abusive	.651	Soils though trained	.595
Tantrums	.643	Preoccupied with trivial items (e.g., string, twig)	.569
Irritable	.634	Urinates though trained	.559
Kicks, hits	.603	Hits self	.527
Noisy	.560	Mouth objects	.513
Impatient	.556	Hums, grunts etc.	.487
Bossy	.546	No sense of danger	.476
Atttention seeking	.545	Repetitive actions (e.g., switches lights)	.459
Mood changes	.543	Loner	.458
Jealous	.536	Runs away	.441
Stubborn	.518	Screams	.414
Throws, breaks	.504	Masturbates in public	.395
Whines	.480	Overactive	.385
Impulsive	.438	Bangs head	.382
Tense	.398	Clothing idiosyncrasies	.370
Excited	.385	Throws, breaks	.362
Hits self	.354	Wanders	.361
Suicidal talk	.354	Flicks, twirls, etc.	.350
Lies	.350		

Communication Disturbance subscale

Echolalia	.623
Perseveration	.612
Talks to self	.594
Thoughts jumbled	.564
Talks in whispers, etc.	.539
Elated for no reason	.456
Pronoun reversal	.424
Talks fast	.395
Delusion	.355

Anxiety subscale

Separation anxiety	.637
Distressed if alone	.571
Fears, phobias	.528
Cries easily	.490
Upset over small changes	.478
Nightmares	.451
Ignores sounds, distressed by sounds	.397
Tense	.368
Shy	.361
Food fads	.351
Irritable	.350

Autistic Relating subscale

Doesn't show affection	.698
Aloof	.586
Doesn't respond to others feelings	.540
Unhappy	.507
Avoids eye contact	.507
Underactive	.475
Facial tics, grimaces	.445
Repeated movements (e.g., handflapping)	.428

Antisocial subscale

Lies	.586
Steals	.573
Hides	.545
Lights fires	.388

[a]The terms used are summary terms, not the actual wording of the items in the checklist.

Table IV. Interrater Agreement and Test–Retest Reliability of Subscale Scores[a]

Subscale	% of Variance	Internal consistency (Cronbach's alpha)	Parent agreement		Teachers/aides agreement		Test–retest	
			ICC	99% CI	ICC	99% CI	ICC	99% CI
Disruptive	15.6	.905	.78	.56–.90	.68	.47–.81	.84	.73–.91
Self-absorbed	6.2	.855	.79	.58–.90	.74	.60–.81	.87	.78–.93
Language deviance	3.6	.806	.75	.50–.88	.62	.42–.76	.76	.61–.86
Anxiety relating	3.0	.762	.80	.61–.90	.66	.49–.78	.77	.62–.86
Autistic relating	2.6	.732	.78	.56–.90	.48	.28–.65	.70	.52–.82
Antisocial	2.3	.670	.79	.60–.90	.30	.05–.50	.51	.27–.70

[a] ICC = Intraclass correlation; CI = confidence intervals.

We have investigated the relationship of scores on the Autistic relating subscale of the DBC with scores on instruments specifically designed to assess the syndrome of autism (Tonge et al., in preparation). One hundred twenty children who met DSM-III-R criteria for Autistic Disorder were assessed with the DBC, the Autism Behavior Checklist (ABC; Krug et al., 1980), and the Autistic Descriptors Checklist (ADC; Friedman, Wolf-Schein, Cohen, & Fisch, 1985). There was a correlation of $r = .74$, $p < .0005$ and $r = .62$, $p < .0005$ between the Autistic Relating subscale and the ABC and ADC respectively. That is, in children with autistic disorder, the DBC Autistic Relating subscale demonstrates concurrent validity in assessing severity of the disorder. The criterion group validity of the same subscale of the DBC was also examined to determine its capacity to discriminate 118 autistic and 156 nonautistic subjects, diagnosed by DSM-III-R criteria. The difference in scores as measured by ANOVA was $F = 43.4$, $p < .0001$.

The Antisocial behaviors are more sophisticated disruptive behaviors that occur with increased prevalence as in mild MR, compared with lower IQ groups (Einfeld, 1993). There is also a significantly lower mean score on this subscale in young people with fragile X syndrome than in age-, sex-, and IQ-matched controls, $F = 9.99$, $p < .002$ (Einfeld, Tonge, & Florio, 1994). The relationship of the subscales to genetic syndromes provides evidence of external validity for these subscales.

The subscales are similar to the factors identified by Aman (1991a) in his review of the factor structure of a variety of instruments used to assess persons with mental retardation, lending the subscales what Aman described as "factorial validity."

Reliability Studies

The reliability of the total scale and each of the subscales was assessed by a series of studies of interrater and test–test reliability. Separate trials of interrater agreement were conducted for mothers and fathers, for nurses in large residential institutions, and for teachers and teachers' aides. Reliability was assessed for the TBPS and for each of the subscales.

Interrater Agreement

The mothers and fathers of intellectually handicapped children attending developmental assessment and child psychiatry clinics each independently completed a questionnaire. In one case, a mother and grandmother who lived in the same house completed the questionnaire. In 24 of the 42 pairs, the clinic nurse supervised the independent completion of questionnaires.

Two nurses caring for disabled adolescents in large residential institutions completed the DBC. In each case, the two nurses had known the person well for at least 6 months and both had cared for the handicapped individual during the day.

Interrater reliability of the Teachers Version was assessed by asking teachers and teachers' aides in special schools and special classes in ordinary schools to independently complete the Teachers Version of the DBC. The teachers and aides were those who knew the pupils well for at least 2 months.

Teacher-parent agreement was assessed by asking one parent (usually the mother) and the teacher of 59 subjects to complete the DBC independently, within 2 weeks of each other. The numbers of subjects who participated in each of these studies and their characteristics are summarized in Table V.

There was good interrater agreement on the total score between the parents, between nurses, and between teachers and aides as assessed by the intraclass correlation (Bartko & Carpenter 1976, Shrout & Fleiss, 1979) (see Table VI). The exception was the poor agreement between parents and teachers, a finding reported in the Isle of Wight Studies (Rutter et al., 1970) and by Achenbach, McConaughy, and Howell (1987, quoted by Aman, 1991a) who found a mean correlation of .27 between parents and teachers in assessments of child psychopathology. The poor agreement in this case probably arises because children and adolescents behave differently at home and at school.

There was also good agreement between the parents ratings on each of the subscale scores (see Table IV above). The agreement between the ratings of teachers and teachers aides was acceptable for the first four of these scales but poor for Autistic Relating and Antisocial Behavior.

Test–Retest Reliability

The test–retest reliability of the total scale score was assessed by asking 63 pairs of parents, residential care workers, and nurses to complete the questionnaire 2 weeks apart. This was a period that might reduce "practice effects" but sufficiently short so as to minimize the possibility that the child's behavior had changed. The test–retest reliability of the Teachers Version was measured by comparing total behavior problem and subscale scores on checklists completed 2 weeks apart by teachers and teachers aides for 13 moderately and severely intellectually handicapped pupils ages 12–17 years. The test–retest reliability coefficient of .83 of the instrument was satisfactory.

Table V. Sample Characteristics for Reliability Studies

Study	% Male	Age	(SD)	Mental retardation (%)		
				Mild	Moderate	Severe
Parent–Parent (DBC-P) TBPS	52.4	11.2	(4.9)	31	45	24
Nurse–Nurse (DBC-P) TBPS	50	16.4	(4.1)	6	41	53
Teacher–Aide TBPS	56	13.4	(3.9)	25	41	34
Teacher–Parent TBPS	59	12.9	(3.7)	49	24	26
Test–Retest (DBC-P) TBPS	55	12.4	(3.4)	7.5	49	43.5
Test–Retest (Teachers) TBPS	54	10.7	(4.2)	—	62	38

Table VI. Reliability Data for Developmental Behavior Checklist[a]

Study	n	% Male	Age	(SD)	Mental retardation			ICC	99% CI
					Mild	Moderate	Severe		
Parent–Parent (DBC-P) TBPS	42	52.4	11.2	(4.9)	31	45	24	.80	.59–.90
Nurse–Nurse (DBC-P) TBPS	32	50	16.4	(4.1)	6	41	53	.83	.68–.92
Teacher–Aide TBPS	110	56	13.4	(3.9)	25	41	34	.60	.42–.74
Teacher–Parent TBPS	51	59	12.9	(3.7)	49	24	26	.05	.31–.41
Test–Retest (DBC-P) TBPS	63	55	12.4	(3.4)	7.5	49	43.5	.83	.69–.90

[a]ICC = intraclass correlation; CI = confidence intervals.

Validity Studies

The content validity of the DBC was assured by the method of item derivation and refinement which has already been described. Evidence of its construct validity has been provided by the results of the principal components analysis of its items. Concurrent validity was assessed by a series of studies of agreement between parents' ratings of symptoms and psychiatrists' ratings of severity of behavior disorder; correlations between scores on other instruments assessing behavioral disorders that require professional knowledge; and correlations between teachers' ratings of behavior disorder and their ratings of the degree of difficulty in teaching the child.

Parent–Psychiatrist Agreement

The first study was of the relationship between ratings of behavior disorders by psychiatrists and DBC scores obtained from parents. Seventy individuals on whom checklists were completed were assessed by two of three child psychiatrists and one clinical psychologist experienced with behavior problems in children with mental retardation (S. E. with either B. T., H. M., or T. F.) who made an overall rating of the severity of psychopathology.

The subjects were children and adolescents attending developmental assessment clinics who had been rated by clinic staff as both behavior disturbed ($n = 38$) and nondisturbed ($n = 32$). The proportion of males was 60%, the mean age was 11.2 years, and the prevalence of mild, moderate, and severe retardation were 37, 29, and 4%, respectively. The professional assessors who were blind to the clinic's assessment provide an overall rating on three scales quantifying the definition of behavioral or emotional disturbance: abnormality or quantitative or qualitative deviance; distress to person or carers; and impairment to adaptive functioning beyond that resulting from developmental disability itself. Each subscale was scored on a 0, 1, 2 rating. These three ratings were summed to produce an overall measure of severity. The Pearson product-moment correlation between the DBC total score and this rating was .81 ($p < .001$)

Correlations with Other Scales

Total scores on the DBC completed by lay informants were correlated with the scores obtained on other rating instruments used for assessing psychopathology in children with MR, all of which must be completed by mental health professionals. These were the Maladaptive Behavior Scale

of the American Association of Mental Deficiency, the Adaptive Behaviors Scales School Edition (Lambert & Windmiller, 1981), and the total score on the Problem Behavior Section of the Scales of Independent Behaviors (Bruininks et al., 1984).

This study comprised 17 male and 23 female subjects: 6 were 0–5 years old, 15 were 6–12 years old, and 19 were 13–20 years old. One subject had an IQ in the borderline range, 9 were mildly intellectually handicapped, and 16 subjects were moderately, and 14 severely handicapped.

There were strong positive correlations between the DBC and each of these scales. The correlations were: .86 between DBC and the Adaptive Behavior Scale and .72 between the DBC and the Scales of Independent Behavior ($p < .001$ in each case).

Prediction of Psychiatric Caseness

The criterion group validity of the DBC was assessed by measuring how well the total score distinguished between individuals who had been classified as "cases" or "noncases" by two experienced psychiatrists. The 70 individuals previously described in the study of parent–psychiatrist agreement were rated as either a definite psychiatric case or not, using the 6-point scale described above. Those rated as 0, 1, 2, and 3 were classified as noncases, and those rated as 4, 5, and 6 were classified as cases.

A Receiver Operating Characteristics (ROC) analysis was conducted to provide a quantitative measure of the ability of the DBC total score to distinguish between cases and noncases over the full range of scores (Fombonne, 1991). At any given cutoff score on the instrument, a 2 × 2 contingency table can be constructed that summarized the accuracy with which the instrument correctly classifies the subject as a true positive, or a true negative, or incorrectly classifies them as a false positive or a false negative.

The ROC curve is a plot of the true positive rate against the false positive rate for each of the family of contingency tables generated by placing the cutoff for defining a case successively at each total scale score. The accuracy with which the instrument makes these classifications is indicated by the area under the ROC curve, which is scaled to have an upper limit of 1.0 for perfect discrimination and a lower limit of 0.50 for chance discrimination.

The area under the ROC curve (estimated by the trapezoid method) was 92% which indicates that the DBC had very good specificity and sensitivity in distinguishing definite psychiatric cases from noncases in this

sample. This figure is superior to that obtained by Fombonne (1991) in a study of the CBCL (Achenbach & Edelbrock, 1983) administered to French children in which the area under the curve was 78% for boys and 82% for girls. The point where the ROC equals 45° is the optimal cutoff. This is a score of 46 on the DBC.

Standardization Norms

The standardization norms for the DBC were derived from a multicenter epidemiological prevalence study of psychopathology in individuals with MR. The study surveyed 171,000 children and adolescents in five census regions of New South Wales and Victoria. Every person 4–18 years old of IQ < 70 identified within this population was included in the study. The number of persons whose parents or primary carers completed the DBC was 454. This was 80.2% of those in the IQ < 50 group identified in the census regions. The participation rate of the IQ > 50 group was lower and there may be some bias in this group towards those with other additional complications of MR such as epilepsy or behavior problems. This difference in ascertainment of subjects with mild MR is a feature of epidemiological samples of MR populations because they tend to blend in with the rest of the population and are identified less completely (Kirman, 1979). The regions are both urban and rural, and have a socioeconomic level which is close to the average for the two States. The study provides percentile distributions of scores at the three levels described above, namely, severity of overall behavior disturbance, disturbance in particular dimensions or syndromes, and disturbance of particular individual behavior or symptoms. Overall levels of behavioral and emotional disturbance were similar for the mild, moderate, and severe MR groups, but were lower in those with profound MR. Full details of the methods and results of this epidemiological study are provided in Einfeld (1992) and Einfeld and Tonge (1993).

The DBC is scored either by hand on the checklist and a scoresheet, or with a computer software program.

APPLICATIONS

Clinical

As behavior disturbance is perhaps the most common complication of mental retardation, the DBC may be used as a routine comprehensive record in developmental pediatric clinics, as suggested by Rowitz (1986). This would allow reliable retrospective behavioral data to be available to

the clinician, and allow natural history and response to intervention to be measured. The DBC may also be used to systematically elicit symptoms in clinical assessments. Cox and Rutter (1985) demonstrated the advantages of structured over unstructured elicitation of symptoms. The norms for the DBC allow the individual patient to be compared in each of the behavioral dimensions measured by the subscales.

Service Planning

The DBC may be utilized to assist in allocation of mental health services to those facilities, regions, or population subgroups with high levels of disturbance. For example, education authorities in the State of New South Wales have used the DBC to assess which school classes for children with MR required extra teaching staff because of high levels of behavior disturbance.

When used as a screening device, the appropriate cutoff point can be determined according to the sensitivity and specificity requirement of the application. Where maximum sensitivity is required, the highest score with 100% sensitivity would be used (i.e., 34). Where maximum specificity is indicated the lowest score with 100% specificity would be chosen (i.e., 68).

Research

In research studies, items, subscales, and total scores can be assessed for their relationship with other variables such as age, sex, IQ, family functioning, or context of disturbance, such as school, home, or workshop.

Individuals and particular subgroups of children and adolescents with MR can be compared against the norms. For example, the DBC has been used to assess psychopathology in children with Prader-Willi syndrome (Hartog, 1992), Williams syndrome (Einfeld, 1992), and fragile X syndrome (Einfeld, Tonge, & Florio, 1994). Other studies are applying the checklist to assess the relation between parental mental state, family function, and behavior disturbance in children with MR.

CONCLUSION

On the data presented the DBC provides a reliable and valid measure of lay informant's ratings of behavioral and emotional disturbances among children and adolescents with mental retardation. The content validity of the items was ensured by the method of selecting and refining its items. A principal components analysis produced six clinically meaningful and factorially valid

subscales which had high internal consistency. Interrater and test-retest reliability of the total score and the six subscales was shown to be satisfactory, with the exception of agreement between parents' and teachers' ratings. There was also good evidence on the validity of the total score. Concurrent validity of total score was shown by correlations between it and psychiatrists' ratings, and scores on other valid instruments for assessing behavioral disorders in children and adolescents with MR which require professional training for their completion. The DBC total score also predicted psychiatrists' classification of subjects with MR as either psychiatric cases or noncases, with a high discriminating power as indicated by the area under the ROC curve.

The current version of the DBC is not necessarily at its maximal state of refinement. Further experience may suggest that some items need to be modified or discarded and new ones added. The correlates of total and subscale scores warrant further study, thus extending the scales' external validity. The DBC should not be regarded as an alternative to assessment by an expert clinician. Nevertheless, when used appropriately, the DBC has the potential to improve existing clinical practice and to advance clinical knowledge about the assessment, diagnosis, management, and prognosis of children and adolescents with MR with behavioral and emotional disorders. The available evidence is sufficient to support its use in research into the prevalence of psychopathology in the population of children and adolescents with mental retardation, and research into its usefulness in clinical practice.

REFERENCES

Achenbach, T. M., & Brown, J. S. (1991). *Bibliography of published studies using the Child Behavior Checklist and related materials.* Burlington; Vermont, University Associates in Psychiatry.

Achenbach, T. M., & Edelbrock, C. (1983). *Manual for the Child Behavior Checklist and Revised Child Behavior Profile.* Burlington; Vermont University Associates in Psychiatry.

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

Aman, M. G. (1991a). *Assessing psychopathology and behavior problems in persons with mental retardation. A review of available instruments* (DHSs Publication No. ADM 91-1712). Rockville, MD: U.S. Department of Health and Human Services.

Aman, M. G. (1991b). Review and evaluation of instruments for assessing emotional and behavioral disorders. *Australian and New Zealand Journal of Developmental Disorders, 17,* 127-145.

Aman, M. G., & Singh, N. N. (1987). Psychometric characteristics of the Aberrant Behavior Checklist. *American Journal of Mental Deficiency, 89,* 492-502.

Aman, M. G., & Singh, N. N. (1988). Patterns of drug use, methodological considerations, measurement techniques, and future trends. In *Psychopharmacology of Developmental Disabilities.* New York: Springer-Verlag.

Ballinger, B. R., & Reid, A. H. (1977). Psychiatric disorder in an adult training centre and a hospital for the mentally handicapped. *Psychological Medicine, 7,* 525-528.

Bartko, J. L., & Carpenter, W. T., Jr. (1976). On the methods and theory of reliability. *The Journal of Nervous and Mental Disease, 163,* 305-317.

Bruininks, R. H., Hill, B. K., & Morreau, L. E. (1988). Prevalence and implications of maladaptive behaviors and dual diagnosis in residential and other service programs. In J. A. F. J. Menolascino, M. H. Albarelli, & V. C. Gray (Eds.), *Mental retardation and mental health.* New York: Springer-Verlag.

Bruininks, R. H., Woodcock, R. W., Weatherman, R. F., & Hill, B. K. (1984). *Scales of independent behavior,* Woodcock-Johnson, Psycho-Educational Measures, Part 4. DLM.

Corbett, J. A. (1979). Psychiatric morbidity and mental retardation. In F. E. James & R. P. Snaith (Eds.), *Psychiatric illness and mental handicap.* London: Gaskell Press.

Cox, A., & Rutter, M. (1985). Diagnostic appraisal and interviewing. In M. Rutter & L. Hersov (Eds.), *Child and adolescent psychiatry modern approaches* (2nd ed., pp. 233-247). Oxford, U.K.: Blackwell.

Einfeld, S. L. (1992). *Behavioral and emotional problems of children and adolescents with mental retardation: Assessment and prevalence in the community and genetic mental retardation syndromes.* A Thesis and Supporting Publications for the degree of Doctor of Medicine, University of Sydney.

Einfeld, S. L., & Aman, M. (1995). Issues in the taxonomy of psychopathology in mental retardation. *Journal of Autism and Developmental Disorders, 25,* 143-167.

Einfeld, S. L., & Tonge, B. J. (1991). Psychometric and clinical assessment of psychopathology in developmentally disabled children. *Australian and New Zealand Journal of Developmental Disabilities, 17,* 147-154.

Einfeld, S. L., & Tonge, B. J. (1993). Population prevalence of behavioral of emotional disturbance in children and adolescents with mental retardation. 2. Epidemiological findings.

Enfield, S. L., Tonge, B. J., & Florio, T. (1994). Behavioral and emotional disorders in fragile X syndrome. *American Journal of Medical Genetics, 51,* 386-391.

Fombonne, E. (1991). The use of questionnaires in child psychiatry research: Measuring their performance and choosing an optimal cut off. *Journal of Child Psychology and Psychiatry, 32,* 677-693.

Fraser, W. I., & Rao, J. M. (1991). Recent studies of mentally handicapped young people's behavior. *Journal of Child Psychology and Psychiatry, 32,* 1, 79-108.

Friedman, E., Wolf-Schein, E. G., Cohen, I. L., & Fish, G. S. (1985). *Autistic Descriptors Checklist (ADC), parent interview edition and scoring guidelines.* Staten Island: New York State Institute for Basic Research in Developmental Disabilities.

Haracopos, D., & Kelstrup, A. (1978). Psychotic behavior in children in institutions for the mentally retarded in Denmark. *Journal of Autism and Childhood Schizophrenia, 8,* 1.

Hartog, S. (1992). *Behavioral characteristics of children with Prader-Willi syndrome.* Masters of Clinical Psychology Thesis, Macquarie University, Sydney.

Heber, R. (1961). Manual on terminology and classification in mental retardation [Monograph]. *American Journal of Mental Deficiency, 66* (Suppl.).

Hepperlin, C. M., Stewart, G. W., & Rey, J. M. (1990). Extraction of depression scores in adolescents from a general purpose behavior checklist. *Journal of Affective Disorders, 18,* 105-112.

Jacobson, J. W. (1990). Assessing the prevalence of psychiatric disorders in a developmentally disabled population. In E. Dibble & D. B. Gray (Eds.), *Assessment of behavior problems in persons with mental retardation living in the community.* Rockville, MD: National Institutes of Mental Health.

Karno, M., & Norquist, G. S. (1989). In H. I. Kaplan & B. J. Sadock (Eds.), *Comprehensive textbook of psychiatry,* Vol. 4. Baltimore: William and Wilkins.

Kirman, B. (1979). Mental illness and mental handicap: Reflections on diagnostic problems and prevalence. In F. E. James & R. E. Snaith, (Eds.), *Psychiatric illness and mental handicap. London: Gaskell.*

Krug, D. A., Arick, J. R., & Almond, P. J. (1980). Behavior checklist for identifying severely handicapped individuals with high levels of autistic behavior. *Journal of Child Psychology and Psychiatry, 21,* 221-229.

Lambert, N., & Windmiller, M. (1981). *AAMD Adaptive Behavior Scale school edition.* Monterey, CA: McGraw-Hill.

Lindsay, W. R., & Mitchie, A. M. (1988). Adaptation of the Zung Self-Rating Anxiety Scale for people with a mental handicap. *Journal of Mental Deficiency Research, 32,* 485-490.

Menolascino, F. J. (1988). Mental illness in the mentally retarded: Diagnostic and treatment issues. In J. A. Stark, F. J. Menolascino, M. H. Albarelli, & V. C. Gray (Eds.), *Mental retardation and mental health.* New York: Springer-Verlag.

Myers, B. A. (1987). Psychiatric problems in adolescents with developmental disabilities. *Journal of the American Academy of Child and Adolescent Psychiatry, 26,* 74-79.

Nihira, K., Foster, R., Shellhaas, M., & Leland, H. (1975). *AAMD Behavior Scale Manual* (rev.), Washington, DC: American Association on Mental Deficiency.

Offord, D. R., Boyle, M. H. Szatmari, P., Rae-Grant, N. I., Links, P. S., Cadman, D. T., Byles, J. A., Crawford, J. W., Blum, H. M., Byrne, C., Thomas, H., & Woodard, C. A. (1987). Ontario Child Healthy Study: II. Six-month prevalence of disorder and rates of service utilization. *Archives of General Psychiatry, 44,* 832-836.

O'Grady, K. E. (1982). Measures of explained variance: cautions and limitations. *Psychological Bulletin, 92,* 766-777.

Parmenter, T. R. (1988). An analysis of Australian mental health services for people with mental retardation. *Australian and New Zealand Journal of Developmental Disorders, 14,* 19-14.

Pfaffenberger, B. (1992). *Using Word 5 for the Mac.* Carmel, IN: Que.

Quine, L., & Pahl, J. (1989). Stress and coping in families caring for a child with severe mental handicap: A longitudinal study. Canterbury, England: University of Kent.

Reid, A. H. (1983, July). Psychiatry of mental handicap: A review. *Journal of the Royal Society of Medicine, 76* 587-592.

Reid, A. H., Ballinger, B. R., & Heather, B. B. (1978). Behavioral syndromes identified in a sample of 100 severely and profoundly retarded adults. *Psychological Medicine, 8,* 399-412.

Rojahn, J., & Helsel, W. J. (1991). The Aberrant Behavior Checklist in children and adolescents with dual diagnosis. *Journal of Autism and Development Disorders, 21,* 17-28.

Rowitz, L. (1986). Editorial, A minimum data set for mental retardation. *Mental Retardation, 24,* 195-196.

Russell, O. (1985). *Mental handicap (Current reviews in psychiatry).* Edinburgh, U.K.: Churchill Livingstone.

Rutter, M. (1967). A Children's behavior questionnaire for completion by teachers: preliminary findings. *Journal of Child Psychology and Psychiatry, 8,* 1-11.

Rutter, M., Tizard, J., & Whitmore, K. (1970). *Education, health and behavior.* London, Longman.

Scott, K. G., Swales, T. P., & Danhour, K. R. (1990). Critical issues in the assessment of childhood psychopathology: An overview of the current state of the art. In E. Dibble & D. B. Gray (Eds.), *Assessment of behavior problems in persons with mental retardation living in the community.* Rockville, MD: National Institute of Mental Health.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Use in measuring rater reliability. *Psychological Bulletin, 86,* 420-428.

Sovner, R. (1986). Limiting factors in the use of DSM-III criteria with mentally retarded persons. *Psychopharmacology Bulletin, 22,* 4.

Sparrow, S. C., Balla, D. A., & Cichetti, D. (1984). *Vineland Adaptive Behavior Scales.* Circle Pines, MN: American Guidance Service.

Spivack, G., & Spotts, J. (1966). *Devereux Child Behavior Rating Scale manual.* Devon, P. A.: Devereux Foundation Press.

Streiner, D. L., & Norman, G. R. (1989). *Health measurement scales.* Oxford, U.K.: Oxford Medical Publications.

Sturmey, P. (1933). The use of DSM and ICD Diagnostic criteria in people with mental retardation. A review of empirical studies. *Journal of Mental and Nervous Disease, 181,* 38-41.

Sturmey, P., Reed, J., & Corbett (1991). Psychometric assessment of psychiatric disorders in people with learning difficulties (mental handicap): A review of measures, *Psychological Medicine, 21,* 143-155.

Wing, L. (1982). *Schedule of handicaps, behaviors and skills.* Unpublished manual. London: Institute of Psychiatry.

Zarfas, D. E. (1988). Mental health systems for people with mental retardation: A Canadian perspective. *Australian and New Zealand Journal of Developmental Disability, 14,* 3-9.