

of
ve
m

Part Five

Technical appendices

Appendix 1: Evaluative measures used

This appendix contains brief details of those main evaluative measures that have not been fully described in Chapter 3, on method. For the sake of completeness we list all the main measures.

Screen measures

Multiple criterion screen techniques were used to identify children with all degrees and types of maladjustment, including those who were 'at risk'. The nature of our screens reflected our concept of maladjustment, and the fact that we were particularly interested in its behavioural and social aspects. Our screen techniques gave rise to a wider spectrum of maladjustment than occurs with a screen based on a single scale and was therefore representative of psychiatric disturbance usually found in clinical practice in a wide range of settings. A full account of our screen techniques is provided elsewhere (Kolvin *et al.* 1977; Macmillan *et al.* 1980; Nicol *et al.* 1981).

RUTTER TEACHER SCALE B2 (JUNIORS AND SENIORS)

Test-re-test reliability for total scores over a three-month interval was reported as 0.89. Inter-rater reliability of 0.72 was reported, but this was over a two- to three-month period (Rutter 1967).

With regard to validity it was reported that total scores on a sample of normal nine- to thirteen-year-old children were significantly lower than scores of a group of children attending a psychiatric clinic.

The B2 scale is a slight revision of the B scale, differing in the wording of a few items. Our use of the former is discussed further in Chapter 3.

336 Help Starts Here

SOCIOMETRY

See Chapter 3.

YOUNG GROUP READING TEST (JUNIORS)

See Chapter 3.

JUNIOR EYSENCK PERSONALITY INVENTORY (JEPI) NEUROTICISM (SENIORS)

A detailed account of this measure has been provided in the evaluation of the results (Chapter 9) as one of its scales (extroversion/introversion) was used as an additional measure.

The JEPI was constructed by Eysenck (1965). It consists of sixty items which generate two dimensions of personality: neuroticism and introversion/extroversion. There is also a lie scale. There is no psychoticism dimension as in the Eysenck personality questionnaire (Eysenck and Eysenck 1975).

We used the inventory for the senior group who were eleven to twelve years of age. At this age the split-half reliability co-efficient of the neuroticism scale is 0.86, of the introversion/extroversion scale it is 0.75, and of the lie scale it is 0.72.

Principal component analyses were carried out for children of different ages and Eysenck concluded from the results that neuroticism can be measured adequately at all the ages covered by the inventory, i.e. seven to sixteen years, but that introversion/extroversion does not clearly emerge as a dimension of personality measurable by questionnaire until the age of nine or ten. Our senior group, on which this inventory was used, were all older than this.

THE RUTTER A SCALE QUESTIONNAIRE (JUNIORS AND SENIORS)

This questionnaire was developed for completion by parents as an instrument to identify psychiatric disorder on a population basis (Rutter, Tizard, and Whitmore 1970). The test-re-test reliability was reported as 0.74. Rutter reported that the scale discriminated well between normal and clinic populations, 67 to 71 per cent of a clinic population achieving a score above cut-off as opposed to 8 to 15 per cent of a normal population sample. The questionnaire yields conduct and neurotic sub-scales composed of items that have been found to discriminate significantly between the two groups of disorders. For the purposes of the present study the conduct and neurotic sub-scales were lengthened. This was done by our including all those items that discriminated each sub-group from the control in the Isle of Wight results (Rutter, Tizard, and Whitmore 1970). The reason for lengthening the scale was to increase its stability. The lengthened

scales
scale a

DEVE
(DESI

Spiva
behav
the cl
that s
specif
distur
study
norma

The
factor
simila
sex),
are th

Non
grade
test re
factor
reliab
range
Zawe
both l

As
the e
teach
with
emoti

It is
additi
provi
study

BARK
These
Natio
of str
and 1

The
interc

Appendix 1: Evaluative measures used 337

scales consisted of items 12, 16, 27, 28, 37, 41, and 42 for the antisocial scale and items 8, 9, 13, 20, 23, 30, and 39 for the neurotic scale.

DEVEREUX ELEMENTARY SCHOOL BEHAVIOUR RATING SCALE (DESB) (JUNIORS AND SENIORS)

Spivack and Swift (1967) designed the DESB to measure overt behaviours that reflect a child's overall adaptation to the demands of the classroom setting and that may affect his or her achievement in that setting. Since their intention was to measure behaviours of specific relevance to the classroom, and not behaviours relating to disturbance in a psychiatric context, the initial pool of 111 items for study was derived from a series of discussions with seventy-two normal- and special-class teachers.

The DESB consists of forty-four items, defining eleven behavioural factors, the items being selected on the basis of factor loadings, similar patterns of correlation with other variables (such as IQ and sex), and significant correlation with academic achievement. There are three additional items that do not contribute to the eleven factors.

Norms are based on ratings by thirty-two kindergarten to sixth-grade teachers of the behaviour of 809 children in the USA. Test-retest reliabilities over a one-week period range from 0.85 to 0.91 for the factors and from 0.71 to 0.80 for the three additional items. Inter-rater reliability, based on forty pairs of ratings in a normal classroom, ranges from 0.62 to 0.77, with a mean of 0.70. Schaefer, Baker, and Zawel (1975) also reported data indicating satisfactory consistency both between raters and over time.

As regards validity, Spivack and Swift (1973) reported that each of the eleven factors had been shown to correlate significantly with teacher grades, after the influence of IQ had been partialled out, both with normal American and French children and with groups of emotionally disturbed children.

It is to be noted that we used both the eleven factors and the three additional items and referred to all fourteen as items. An account is provided in Appendix 2 of the factor analysis undertaken in this study.

BARKER LUNN CHILDREN'S ATTITUDE SCALES - S7 (SENIORS)

These attitude scales were developed in the context of a study by the National Foundation for Educational Research (NFER) into the effects of streaming and non-streaming in junior schools (Barker Lunn 1967 and 1969).

The questionnaire contains ten attitude scales: attitude to school; interest in school work; importance of doing well; attitude to class;

338 Help Starts Here

'other' image of class; conforming versus non-conforming pupil; relationship with teacher; anxiety about school work; social adjustment; and self-image. Each scale is made up of six to ten statements made by children during group discussions and selected by factor analysis and scalogram analysis.

Some 2300 third- and fourth-year junior schoolchildren in twenty-eight schools completed the final form of the questionnaire. Inter-correlation of the scales showed two clusters: one dealing with attitudes towards school and school work and the other with social relationships and the personality of the pupil. The internal consistency of the scales was determined by Cronbach's Alpha-coefficient, yielding a range of 0.69 to 0.90 with a mean of 0.81. A large number of correlations are reported with other measures, such as teacher and parent ratings of ability, sociometric data, achievement scores, and interest scores.

The expectation of a relationship between school performance and attitude was borne out. All the scales correlated significantly with test scores in English, problem arithmetic (all except conforming/non-conforming), essays, verbal reasoning, and non-verbal reasoning. A similar pattern was evident for teacher ratings and also (with the exception of conforming/non-conforming) with parent ratings. The social adjustment scale correlated 0.21 with sociometric status. Interest scores correlated most highly with 'attitude to school' and 'interest in school work'.

An account is provided in Appendix 2 of the factor analysis undertaken in the current study.

The Devereux and Barker Lunn results have been presented separately because we considered them to be measures of general functioning and attitudes highly relevant to the classroom setting.

Child behaviour and temperament: based on parental reports (juniors and seniors)

BEHAVIOUR

An inventory, administered as a semi-structured, open-ended interview with mothers, was used to quantify behaviour (Kolvin *et al.* 1975b). It consists of twenty-nine questions with appropriate probes which relate to three of the four scales originally developed - neurotic behaviour (scale A), antisocial behaviour (scale B), and psychosomatic behaviour (scale C). The inter-rater reliability of the original scales were all above 0.90. For the purposes of the study we adopted the above three scales but modified and extended the content so that

it was more clinically relevant and more appropriate to an older population of children. The scores on the above three scales were summed to provide an aggregate behaviour score (scale A + B + C). In addition, we added a narrower dimension of purely somatic disturbance which reflects abdominal pain, headaches, and vomiting.

TEMPERAMENT

An inventory, again administered as a semi-structured, open-ended interview with mothers, was used to measure temperament (Garside *et al.* 1975). This, too, was slightly modified to meet the purposes of the current study. There are twenty-nine questions with appropriate probes which relate to the four dimensions of withdrawal, activity, mood, and irregularity. The inter-rater reliabilities of these dimensions were all above 0.90.

Additional information – cognitive

When deciding which cognitive and achievement measures to employ, in addition to technical considerations, we found that ease of application of tests and, where possible, group application, were important, since such large numbers of children were involved and disruption of schools' normal working routines had to be minimized. Another limiting factor in the selection of measures was the need for a test to span the age range encountered in the three-year period in which assessments were conducted. We included the English Picture Vocabulary Test and Holborn Reading Scale in our battery with the junior children since data from these tests were available to us from the local education authority's routine assessments in the schools.

ENGLISH PICTURE VOCABULARY TEST (EPVT) (JUNIORS)

The English Picture Vocabulary Test (Brimer and Dunn 1962) is described as measuring 'listening vocabulary' or, more generally, verbal comprehension, and is regarded as being functionally and administratively independent of reading skill.

Form 2 of the test was employed in this study. It covers the age range of seven years to eleven years and eleven months, and can be used as either an individual or group test. The test comprises forty items arranged in order of increasing difficulty. A page of pictures together with a spoken word constitutes a test item and the child is required to identify the picture to which the word refers.

The EPVT was derived from the Peabody Picture Vocabulary Test (Dunn 1959), developed in the USA. English standardization of the EPVT (2) was conducted in Wiltshire schools, being administered to

340 Help Starts Here

over 5000 children. Kuder Richardson reliabilities are presented for each year group, and the mean reliability is 0.92.

As a means of measuring concurrent validity, the EPVT (2) was administered to 223 primary schoolchildren aged between eight years and eleven years eleven months, along with the Schonell Comprehension Test (Schonell and Schonell 1963), and an experimental test of 'expressed' vocabulary involving 'written sentence completion within an orally presented sentence context' (Brimer and Dunn 1962:35). A random sample of seventy-eight of these children was also administered the Wechsler Intelligence Scale for Children (WISC) Vocabulary and the Schonell Graded Word Reading Test (GWR). The product-moment correlations with Schonell Comprehension and Expressed Vocabulary were 0.61 and 0.73 respectively, and with WISC Vocabulary and Schonell GWR, 0.76 and 0.80 respectively. It was suggested that these results, especially the correlations with tests purporting to measure vocabulary, provide evidence that EPVT (2) measures a function common to other tests of vocabulary.

The correlation of 0.76 with WISC Vocabulary supported Brimer and Dunn's claims that the EPVT (2) yields a measure indicative of verbal ability. In addition, a correlation of 0.79 with the verbal ability test in the eleven-plus examination was reported for 271 of the standardized sample. Further evidence of the relationship with verbal ability came from a study by Phillips, cited in the test manual, where a correlation of 0.81 with Stanford-Binet Vocabulary was found with a sample of 124 ten- to eleven-year-olds. This sample was reported to have departed from national representativeness in sex distribution.

MORAY HOUSE PICTURE INTELLIGENCE TEST (MELLONE 1948) (JUNIORS)

This is a well established test and, according to the manual, has a reliability of about 0.95. It has been reviewed by Banks and Pringle (in Buros, *The Fifth Mental Measurements Yearbook* (1959)). We used it to test the non-verbal intelligence of junior children at the initial testing. By the time of the midline assessment the age ceiling of the Moray House had been reached and at this assessment, and at the final one, the Cattell Intelligence Test Scale 1 was employed.

CATTELL INTELLIGENCE TEST SCALE 1 (JUNIORS)

This is an established, well standardized, non-verbal test of intelligence, although there is no information regarding its reliability or validity. It has been well reviewed by MacFarlane Smith (in Buros, *The Fifth Mental Measurements Yearbook* (1959)). We used it in the

Appendix 1: Evaluative measures used 341

follow-ups of the junior children. The IQs on this test were rescaled to make them comparable with those of the Moray House Test.

HOLBORN READING SCALE (WATTS 1948) (JUNIORS)

This is an individual test of reading and comprehension and was used by us to test the reading of the junior children. There is no information regarding its reliability, but according to Nisbet (in Buros, *The Fifth Mental Measurements Yearbook*, (1959)) it is 'quite obviously a sound test' (Buros 1959:635).

GENERAL ABILITY TEST (USED BY THE NATIONAL CHILD DEVELOPMENT STUDY (NCDS)) - PREVIOUSLY DEVELOPED BY THE NFER (SENIORS)

This test comprises alternate verbal (forty items) and non-verbal (forty items) with eight practice items. It yields a verbal score, a non-verbal score, and a total score. The task for each verbal item is to discover the principle or concept underlying the grouping of four words and then to supply the missing word - on a multi-choice basis - for a second group of three words, according to the same principle. The task is the same for non-verbal items but here the groupings involve shapes. It is designed for group administration.

At age eleven, test-re-test reliability of 0.94 was reported by Douglas (1964). Douglas found a correlation of 0.93 with the NFER verbal test 8A (eleven-plus selection test) ($n = 74$) and, in addition, reported from his survey data with an eleven-year-old sample correlations of 0.69 with both a mechanical reading test and a vocabulary test, and 0.75 with an arithmetic test.

READING COMPREHENSION TEST (DEVELOPED BY NFER) (SENIORS)

This is a thirty-five-item test of reading comprehension, constructed as a parallel test to the Watts-Vernon (Start and Wells 1972), so that for each item in the Watts-Vernon there is an item of comparable facility value in the parallel test. The child is required to choose the correct word from a selection of five given, in order to complete a sentence meaningfully. It is designed for group administration.

A test-re-test correlation of 0.90 was reported by Douglas (1964) for a sample of 124 eight-year-olds, as well as correlations of 0.87 with the mechanical reading test and 0.68 with the vocabulary test employed with this sample.