CHAPTER 4

Behavioural and Temperamental Assessment

If we wished to demonstrate significant differences of behaviour or of temperament, comparable to the inter-group differences demonstrated in the previous chapter, we needed to use methods which could produce comparable results, expressed in terms of a reproducible and reliable score. Only such results could make a useful contribution towards further investigating the possibility that the abnormalities of intra-uterine growth in which we are interested could produce long-term ill-effects as a result of minimal organic brain damage. We think we have been able to use such methods, and will here describe first the methods used and then the results obtained under the separate headings of behaviour and temperament.

Behaviour

We used four types of data to assess this aspect of the children's functioning, which we can call 'what they did'.

Mothers' Reports

A procedure for systematically recording such reports was first introduced by McFarlane *et al.* (1954) and later modified by Grant (1958) to produce a questionnaire in which each item is rated on a five-point scale, ranging from a score of 1 for no abnormality to a score of 5 for an extreme degree of abnormality.

This procedure has been elaborated by Wolff (1967) to obtain from mothers reliable ratings of their children's behaviour by applying a technique which is described as 'focussed interviewing'. The interviewer goes through an inventory of specified questions in a set order so as to elicit from the mother descriptions of what her child does in a series of specified situations; the answers can be rated according to the five-point scale. Questions concerning objective items of behaviour (*e.g.* bed-wetting) are focussed upon the frequency and the severity of the behaviour. Questions concerning relatively subjective items (*e.g.* destructiveness) are brought into focus by a series of enquiries designed to elicit the likelihood of the behaviour appearing in specified situations. The questions are open-ended, however, in that they allow the interviewer to explore the mothers' answers more deeply, if necessary, to achieve a satisfactory rating. For each item the inventory provides a clear definition of every point on the rating scale, from 1 to 5.

There were 38 items in the inventory we administered at the ages of five and seven years, referring to such phenomena as nightmares, soiling, wetting and solitariness. For purposes of description and analysis these items were later grouped into 'dimensions' by two different types of criteria: (*a*) clinical, arrived at by considering what appears to be meaningful in clinical terms in giving rise to dimensions such as sleep disorder or excretory disorder; and (*b*) statistical, arrived at

by a technique of multivariate analysis which is discussed by Kolvin *et al.* (1975). In this chapter we will be reporting the results first in terms of the items in which there was at least one significant difference between our groups at ages of five and seven years, and secondly in terms of the clinical dimensions in which there was at least one significant difference. We have also added together the scores for all the individual behaviour items at each age to give a total behaviour score at each age (with a range of 38 to 190, or 1 to 5 for each of 38 items).

The results at the age of five years are summarised in Table 4.1 and at seven years in Table 4.2. It is immediately obvious that the pattern which we saw in Chapter 3 for the results of the psychometric assessment is repeated here. There are many more significant differences between the very light-for-dates and the control group (to the disadvantage of the former) than there are between the short-gestation and the control group, or between the two extreme abnormal groups. At the age of five years the very light-for-dates group had a significantly worse score than the controls for seven items, and for the total behaviour score. The short-gestation group had no item scores significantly worse than the controls and two scores significantly better than the light-for-dates.

At the age of seven years the very light-for-dates group had a significantly worse score than the controls for five items, for one dimension and for the total behaviour score. The short-gestation group had a significantly worse score than the controls for two items and a significantly better score than the very light-for-dates for one item. The tendency for the scores of the rather light-for-dates to fall between those of the control and the very light-for-dates is less definite than it was in the case of the psychometric tests, and at seven years the intermediate group's score was actually significantly better than that of the control group for the item 'worry about health'.

Teachers' Reports

By the time our children reached the age of seven years, Rutter (1967) had described his Inventory (Scale B) for completion by the teacher and we were able to arrange for this to be completed for all the children still in our study. The inventory consists of 26 items, the answers being rated on a three-point scale (0 to 2). Rutter (1967) has shown that a total score of 9 can be regarded as a meaningful cut-off point between 'normal' and 'abnormal' in what must be regarded as a screening rather than a diagnostic procedure. We have summarised our results in Table 4.3 in terms of the items in which there was at least one significant difference between groups and in terms of the proportion of children in each group whose score was above 9 ('abnormal'). Again there are more significant differences between the control and the very light-for-dates (four) than between the control and the short-gestation group (one).

Examiners' Observations

We mentioned in the previous chapter (see Procedure) that during each testing session the psychologist rated each child's behaviour for shyness, negativism and distractibility on a defined five-point unipolar scale, in which the higher scores represent greater degrees of abnormality. This rating was made at the ages of five, six

and seven years. The psychiatrist who examined the child next, during the same visit to the Unit in nearly all cases, carried out the same rating procedure on the basis of his observations during his neuropsychiatric examinations at the ages of five and six years. All ratings were carried out 'blind', in the sense that the observers knew neither to which of our groups each child belonged, nor each other's ratings. We have

TABLE 4.1

Behaviour: mothers' reports at age five years*

	Random control (a)	*Short gestation (b)*	*Rather light-for-dates (c)*	*Very light-for-dates (d)*	*Significant difference* $p<0.05$	$p<0.01$
Items (out of a total of 38)						
Restless sleep	1.8	2.1	1.8	2.2	a/d	—
Poor appetite	1.4	1.5	1.5	1.9	b/d	a/d
Finickiness	1.5	1.6	1.6	1.8	a/d	—
Wet by night	1.3	1.2	1.4	1.5	a/d, b/d	—
Destructive	1.4	1.5	1.4	1.5	a/d	—
Headache	1.2	1.4	1.3	1.5	a/d	—
Social Isolation	1.1	1.1	1.3	1.2	a/c, a/d	—
Obsessionality	1.3	1.5	1.4	1.4	a/b	—
Speech	1.3	1.5	1.2	1.3	a/b	—
Dimensions (out of a total of 5)						
No significant differences						
Total score, mean	61.72	63.95	62.97	64.11	a/d	—
Standard deviation	7.06	8.97	6.88	7.72		

*Mean abnormality scores on five-point scale (1 to 5) for items with significant differences between groups. Total score is sum of scores for all items, regardless of significant differences.

TABLE 4.2

Behaviour: mothers' reports at age seven years

	Random control (a)	*Short gestation (b)*	*Rather light-for-dates (c)*	*Very light-for-dates (d)*	*Significant difference* $p<0.05$	$p<0.01$
Items (out of a total of 38)						
Dreams	1.4	1.7	1.5	1.8	a/b	a/d
Poor appetite	1.2	1.3	1.2	1.6	—	a/d
Finickiness	1.6	1.7	1.8	2.0	—	a/d
Wet by night	1.2	1.3	1.3	1.5	a/d	—
Reckless	1.3	1.3	1.3	1.7	b/d	a/d
Obedience	1.2	1.3	1.2	1.3	a/d	—
Worry about health	1.6	1.8	1.2	1.4	a/b, b/d	a/c
Headache	1.2	1.4	1.3	1.4	a/b	—
Dimensions (out of a total of 5)						
Psychosomatic	1.6	2.1	1.5	2.4	—	a/d
Total score, mean	54.37	56.16	53.86	57.39	—	a/d
Standard deviation	6.60	8.97	6.51	9.18	—	—

summarised their findings at the ages of five, six and seven years in Table 4.4, in which are shown all items where there was at least one significant difference between groups.

The psychologist's findings show the same tendency as those reported in the previous chapter. Of the nine ratings made, the very light-for-dates group showed significantly more abnormality than the control group in five ratings, the short-gestation group showed more abnormality than the controls in two, and less than both light-for-dates groups in one rating ('negativism' at six years). It is noticeable that there is a general trend towards lower scores (*i.e.* less abnormality) and towards less differences between the groups with successive rounds of testing. The psychiatrist's ratings showed only one significant difference (out of a possible six) in which the short-gestation group showed more abnormality than the controls.

TABLE 4.3

Behaviour: teachers' reports at age seven years*

	Random control (a)	*Short gestation (b)*	*Rather light-for-dates (c)*	*Very light-for-dates (c)*	*Significant difference* $p<0.05$	$p<0.01$
Items						
Fidgety	0.29	0.45	0.35	0.47	a/d	—
Solitary	0.29	0.31	0.32	0.46	a/d	—
Attention-seeking	0.48	0.66	0.53	0.70	a/d	—
Fussy	0.13	0.26	0.23	0.13	a/b	—
Proportion with total score 9 ('abnormal')	15.0%	22.4%	17.0%	30.0%	—	a/d

*Mean abnormality scores on three-point scale (0 to 2)

TABLE 4.4

Behaviour: examiner's observations at ages five, six and seven*

	Random control (a)	*Short gestation (b)*	*Rather light-for-dates (c)*	*Very light-for-dates (d)*	*Significant difference* $p<0.05$	$p<0.01$
Psychologist						
At 5 years: Shyness	1.9	2.2	2.1	2.2	a/b	a/d
Negativism	1.2	1.4	1.3	1.5	a/d	—
Distractibility	1.4	1.5	1.4	1.7	c/d	a/d
At 6 years: Negativism	1.1	1.0	1.2	1.2	a/c, a/d	b/c, b/d
Distractibility	1.1	1.2	1.2	1.3	a/d	—
At 7 years: Distractibility	1.1	1.3	1.2	1.2	—	a/b
Psychiatrist						
At 5 years: Distractibility	1.4	1.8	1.6	1.7	a/d	a/b
Negativism	1.3	1.5	1.4	1.4	a/b	—

*Mean abnormality scores on five-point scale (1 to 5)

Psychiatrist's Assessment at Age Seven Years

This was carried out by a standard interview of the child, modelled on the one used in the Isle of Wight survey (Rutter and Graham 1968, Rutter *et al.* 1970). The Isle of Wight survey specified 21 areas to be covered and categorized the codings precisely but left the exact wording of the questions to the individual psychiatrist. It also included an over-all assessment of psychiatric abnormality, rated on a three-point scale. We increased the areas covered to 48; we also aggregated the items into clinically meaningful dimensions and summarised their scores to give a total maladjustment score.

The results are summarised in Table 4.5. This is the first of all the assessments whose results we have reported so far in which the short-gestation group shows more significant differences from the controls (four items) than does the very light-for-dates (two items). The same relationship between the two extreme abnormal groups is seen in the rating of over-all psychiatric abnormality, though the difference is slight. In the results of the analysis by dimensions, however, the relationship is reversed, there being five significant differences between the very light-for-dates group and the controls, compared with four such differences in the case of the short-gestation group. The total maladjustment scores of the two extreme abnormal groups (in terms of their intra-uterine growth experience) show an approximately equal and very highly significant difference from the control group.

TABLE 4.5

Psychiatric abnormality score at age seven years

	Random control (a)	*Short gestation (b)*	*Rather light-for-dates (c)*	*Very light-for-dates (d)*	*Significant difference* $p<0.05$	$p<0.01$	$p<0.001$
	%	%	%	%			
(a) **Items* [total 33)							
No special friend	11.3	27.6	7.8	10.1	b/d	a/b	—
Overactivity	18.1	50.0	21.0	42.0	—	a/d	a/b
Fidgetiness	46.3	72.4	46.9	72.5	—	—	a/b, a/d
Poor attention/ persistence	17.5	25.9	50.8	56.5	a/b	b/d	a/c, a/d
Over-all psychiatric abnormality	10.2	26.0	10.5	21.7	—	a/b, a/d	—
	Mean scores						
(b) †*Dimensions*							
Sociability	0.9	1.4	1.0	1.0	b/d	a/b	—
Aggression	1.4	1.4	1.1	1.0	a/c, a/d	—	—
Phobic anxiety	2.7	3.5	3.1	3.7	a/d	—	—
Interview anxiety	2.1	3.2	2.8	3.7	a/b	—	a/d
General anxiety**	4.8	6.8	5.9	7.4	a/b	a/d	—
Motor activity	3.3	5.1	3.5	4.9	—	c/d	a/b, a/d
Total abnormality score	10.4	14.7	11.5	14.3	c/d	—	a/b, a/d

*(a) Proportion showing moderate or severe degree of abnormality in items listed.
†(b) Mean scores for individual dimensions, and total score abnormality.
**General anxiety is the sum of Phobic and Interview anxiety.

Temperament

We used only one type of data to assess this aspect of the children's functioning, which we can call 'how they did things'.

Mothers' Reports

We developed our own inventory, using the same techniques of interviewing and rating as described above, but in this case based on the Wolff Behaviour Inventory (Garside *et al.* 1975). It consisted of 48 items (*e.g.* mood at meal-times, and during play), each rated on a unipolar five-point scale, with a score of 1 for a low level and a score of 5 for the highest level of each item. This Temperament Inventory was completed by interviewing the mothers when the children were five and again when they were seven years old. As in the case of the Behaviour Inventory, we have grouped these items into a number of 'dimensions' for purposes of description and analysis. In this case we have not produced a total score, comparable with the total Behaviour score, because there is no consistent value judgment in the different items and so no consistent direction in the individual scores.

The results at the age of five years are summarised in Table 4.6, showing all the items and dimensions in which there was at least one significant difference between our groups. The same pattern seen in the Behaviour scores is repeated here: the abnormal group which differs most from the controls is the very light-for-dates (in eight items and two dimensions); the short-gestation group differs significantly from the controls in two items, and from the very light-for-dates in only one. Of the eight items in which the score of the very light-for-dates group was significantly worse than that of the controls, five were related to a pattern of high activity — intensity and distractibility. This high activity-intensity problem in the very light-for-dates group is again apparent in the analysis by dimensions, where it is joined by an excess of 'moodiness' in both light-for-dates groups.

TABLE 4.6

Temperament: mothers' reports at age five years*

	Random control (a)	*Short gestation (b)*	*Rather light-for-dates (c)*	*Very light-for-dates (d)*	*Significant difference* $p<0.05$	$p<0.01$
Items (out of a total of 48)						
Play activity	1.8	2.0	2.0	2.1	a/d	—
Dress activity	1.2	1.2	1.3	1.3	a/c, a/d	—
Waking irregularity	3.8	3.9	3.4	4.3	a/d	a/c
Dress intensity	1.1	1.2	1.2	1.2	a/b, a/c, a/d	—
Sulkiness	1.1	1.1	1.1	1.2	a/d	—
Dependency	1.6	1.9	1.7	2.0	a/d	—
Attention impersistence	2.2	2.7	2.4	2.7		a/b, a/d
Distractibility	3.0	3.0	3.1	3.4	a/d, b/d	—
Dimensions (out of a total of 4)						
Activity—intensity	14.3	14.6	14.6	15.5	a/d	—
Moodiness	6.6	6.6	7.1	7.0	a/c, a/d	—

*Mean Abnormality Scores on a five-point scale (1 to 5)

The results at the age of seven years are summarised in a similar way in Table 4.7. This time the very light-for-dates group differs significantly from the control group in nine items (five of which are part of the activity-intensity-distractibility pattern, and four of the irregularity). The short-gestation group differs from the controls in five items and from the very light-for-dates in two, these being in the assertiveness-aggressive pattern, in which the short-gestation group's scores are the lowest of the three. The analysis by dimensions also contains the pattern of high activity-intensity in the very light-for-dates group, but at this age of seven years it is also seen in the short-gestation group, in which there is also an excess of 'moodiness'. This latter is no longer present in the very light-for-dates group but there is now a significantly high score for irregularity, thus completing what we may fairly describe as an over-all hyperkinetic picture for this group.

Discussion

As in the case of the psychometric tests, our ratings of both behaviour and temperament show a clear excess of abnormalities among the children in both extreme abnormal groups as compared with the controls. However, although there is again a larger number of significant abnormalities in the very light-for-dates group as compared with the short-gestation group in the results which we have reported under the heading of 'behaviour', there is no such difference in the results under the heading of 'temperament'. Between the two extreme abnormal groups there are again surprisingly few differences, possibly no more than would be expected to occur by chance when such a large number of ratings has been carried out. The general picture is that of a hyperactive pattern in both the extreme abnormal groups, but particularly in the very light-for-dates. It is noteworthy that in certain items the children in the short-gestation group show significantly less abnormality than either the controls or the very light-for-dates; for instance they show less assertion-dominance and less spontaneous aggression (see Table 4.5).

However, as with the results of the psychological tests, we cannot begin to attribute the differences we have demonstrated directly to the abnormalities of intra-uterine growth themselves until we have allowed for the associated 'confusing' factors. In this case there seems no reason to consider correcting for differences in mean gestational age since they can hardly affect behavioural or temperamental ratings.

The differences between boys and girls are very interesting, however. In Table 4.8 we have summarised the results for the total Behaviour Abnormality scores, based on the mothers' reports (see Tables 4.1 and 4.2). In this table, and in the remainder of this chapter, we have not subdivided the light-for-dates group for the same reasons as in the previous chapter. It is at once apparent that, among the girls, those in the abnormal groups show no significant differences from the controls, whereas, among the boys those in the abnormal groups from the point of view of intra-uterine growth show significantly more behavioural abnormalities than the controls in three of the four possible comparisons at five and seven years. In Table 4.9 we have summarised the results of a similar analysis of the Behaviour Abnormality Scores rated by the psychiatrist's direct observations at the age of seven years and these show similar sex

differences. In this case there are two dimensions in which an abnormal group of girls has a more abnormal score than the controls: both are in the short-gestation group, whose total score is also higher than the controls. But among the boys the abnormal groups have a significantly higher score than the controls in seven out of the 10 comparisons of dimensions and in both the comparisons of total score. Clearly the boys appear to be more vulnerable to the effects of abnormalities of intra-uterine growth.

TABLE 4.7

Temperament: mothers' reports at age seven years

	Random control (a)	*Short gestation (b)*	*Rather light-for-dates (c)*	*Very light-for-dates (d)*	*Significant difference* $p<0.05$	$p<0.01$
Items (out of a total of 48)						
Meals activity	1.8	2.1	1.9	2.2	—	a/d
Dress activity	1.4	1.7	1.4	1.7	—	a/b, a/d
Bowel irregularity	1.0	1.2	1.0	1.2	a/d	a/b
Time irregularity	1.4	1.2	1.5	1.7	a/d, b/d	—
Sleep irregularity	1.5	1.6	1.5	1.8	a/d	—
Mood waking	1.1	1.6	1.3	1.2	—	a/b
Assertion-dominance	1.9	1.7	1.8	2.0	a/b, b/d	—
Spontaneous aggression	1.5	1.3	1.4	1.5	a/b, b/d	—
Dependency	1.3	1.5	1.2	1.4	a/b	—
Malleability, call to meals	1.6	1.8	1.6	1.8	a/d	a/b
Impersistence	1.8	1.9	1.8	2.1	a/d	—
Attention	1.3	1.5	1.6	1.6	a/d	—
Distractibility	1.6	1.8	1.7	1.9	—	a/d
Dimensions (out of a total of 4)						
Activity—intensity	12.5	13.9	12.7	14.2	—	a/b, a/d
Moodiness	7.9	8.8	8.0	8.5	a/b	—
Irregularity	8.9	9.5	8.8	10.4	—	a/d

TABLE 4.8

Total behaviour abnormality score related to sex*

	Random control (a)	*Short gestation (b)*	*Light-for-dates (c)*	*Significant difference* $p<0.05$	$p<0.01$
Age 5 years					
Boys (1)	60.5	64.9	63.9	—	a/b, a/c
Girls (2)	63.1	62.7	63.3	—	—
Age 7 years					
Boys (3)	53.9	57.2	55.8	a/b	—
Girls (4)	54.9	54.6	54.9	—	—
Significant difference $p<0.05$	1/2	—	—		

*Mean scores at five and seven years (based on mothers' reports).

TABLE 4.9

Psychiatric abnormality score and breakdown by 'dimensions' related to sex*

		Random control (a)	*Short gestation (b)*	*Light-for-dates (c)*	*Significant difference* $p<0.05$	$p<0.01$
BOYS						
Dimensions						
Sociability	(1)	1.0	1.5	0.9	a/b, b/c	—
Aggression	(2)	1.7	1.6	1.3	a/c	—
Phobic anxiety	(3)	2.4	3.0	3.2	—	—
Interview anxiety	(4)	2.0	3.7	3.8	a/b	a/c
Motor activity	(5)	3.6	5.8	4.8	a/c	a/b
Total score		10.7	15.6	14.1	a/c	a/b
GIRLS						
Dimensions						
Sociability	(6)	0.8	1.4	1.0	a/b	—
Aggression	(7)	1.0	1.1	0.9	—	—
Phobic anxiety	(8)	3.4	4.3	3.6	—	—
Interview anxiety	(9)	2.2	2.6	2.8	—	—
Motor activity	(10)	3.1	4.2	3.7	a/b	—
Total score		10.2	13.6	12.0	a/b	—
Significant difference						
$p<0.05$		—	4/9	2/7, 5/10		
$p<0.01$		2/7	—	—		

*Mean scores at seven years (rated by psychiatrist using 48-item inventory).

Social class is the other major, associated, 'confusing' factor whose effects we have evaluated in a similar way. The results with respect to the Behaviour Abnormality Score are summarised in Table 4.10. They show that among the controls there is a significant difference—in the direction of less abnormality—in children from classes I and II at the age of five years, but by the age of seven years there is not even a suggestion of such a difference. In the short-gestation group the number of children is small, and although there is an obvious trend towards increasing abnormality as we move from classes I and II to IV and V, the difference is not statistically significant. In the light-for dates group, however, there is a significant increase in the abnormality scores as we move from class III to classes IV and V at both five and seven years of age. It is also noteworthy that the largest differences between the two abnormal groups and the control group, and the only ones which reach statistical significance, are in social classes IV and V.

The simple relationships between social class and the Psychiatric Abnormality score are summarised in Table 4.11. This time the obvious trend in the control group—towards a higher abnormality score as we move from classes I and II to IV and V—is not statistically significant, but the similar trend in the light-for-dates group is very significant as we move from class III to classes IV and V. And the

TABLE 4.10

Total behavioural abnormality score, related to social class*

	Random control (a)	*Short gestation (b)*	*Light-for-dates (c)*	*Significant difference p<0.05*	*p<0.01*
Age 5 years					
I + II (1)	57.9	(59.5)	61.7	—	—
III (2)	62.1	62.6	62.7	—	—
IV + V (3)	62.8	65.7	65.6	—	—
Age 7 years					
I + II (4)	54.1	(53.0)	54.9	—	—
III (5)	54.7	54.2	54.6	—	—
IV + V (6)	53.6	58.5	57.9	—	a/b, a/c
Significant difference p<0.05	1/2, 1/3	—	2/3, 5/6		

*Mean scores at five and seven years (based on mothers' reports).

TABLE 4.11

Psychiatric abnormality score, related to social class*

	Random control (a)	*Short gestation (b)*	*Light-for-dates (c)*	*Significant difference p<0.05*	*p<0.01*
I + II (1)	7.9	(10.0)	10.7	—	—
III (2)	10.4	14.5	11.9	—	a/b
IV + V (3)	11.2	15.5	15.6	a/b, a/c	—
Significant difference p<0.05	—	—	1/3, 2/3		

*Mean scores at seven years (rated by psychiatrist using 48-item inventory).

children in social classes IV and V in both abnormal groups, and those in class III in the short-gestation group, have significantly higher scores than the corresponding classes in the control group. Again it looks as though the abnormalities of intra-uterine growth produce more adverse effects in children from an unfavourable social background than in those from a favourable background.

Our conclusions about the 'confusing' factor of broad social-class differences must be that social class is potentially important and must be allowed for when comparing the behaviour or the temperament of the children in an abnormal group with the control group if there is a significant difference in their social class distributions, particularly if there is an excess of children in classes IV and V. Table 2.3 (p. 12) makes it clear that such a difference definitely is present when the short-gestation group is compared with the controls, but equally definitely is not present in the case of the light-for-dates groups (there is, in fact, an excess of children in social classes I and II in the very light-for-dates group).

The findings reported by Drillien (1964) in children up to five years of age are similar to ours in that there was an excess of behaviour problems and of maladjust-

ment in school among her children of very low birthweight (4lb 9oz (2070g) and under) and in her social grade 4. The fact that our incidence of 'over-all psychiatric abnormality' (see Table 4.5) was much lower than the incidence of 'behaviour problems' which Drillien reported (ranging from 33.3 to 93.4 per cent in different subgroups) may be due to the fact that we used relatively objective and rigidly defined methods of assessment. The proportion of children whom Drillien found to be 'maladjusted' in school, as judged by the Bristol Social Adjustment Guide, ranged from 9.3 to 26.8 per cent, and is remarkably similar to the proportion with an 'abnormal' score of more than 9 in our Teacher's Behaviour Inventory (see Table 4.3).