Part II Experimental Design

CHAPTER 3

Method

3.1 Introduction

The methodology on participants and apparatus is divided into two parts, Stage 1 and Stage 2. These reflect experimentation over two main time periods. Stage 1 concerns a study of multiplication number facts born of initial investigations by Pritchard et al. (1989), the rationale for which is discussed below.

Stage 2 has taken place with new participants at a later date and is concerned with experimentation on addition, subtraction and division, thus rounding off mathematical investigations into the four operations. Knowledge gleaned from Stage 1 has helped to furnish subsequent research with improvements and adjustments.

Sections 3.8 to 3.11 detail the methodology for the four experiments covering each of the mathematical operations. The Multiplication experiment is presented first since it was conducted as the initial experiment. The general layout of each experiment and the method employed is similar. This research addresses eight questions as given in Chapter 2.

The dependent variables are temporal ratio data such as *time* taken to respond to the mathematical questions including where appropriate components in this process and also categorical nominal data on the *accuracy* or otherwise of the participants' answers. The independent variables are the *three age bands* (young, medium-age and old) as well as the *experimental groups* (dyslexics, CA controls – referred to as 'controls' in the Division, Addition and Subtraction experiments, as well as the SA controls – in the Multiplication experiment). In addition there are subordinate independent variables such as *order of trial*, *table number* and *quotient.* There are also sub-groupings of tasks within each mathematical operation, such as multiplication and division tables and addition and subtraction categories that are also included in the experimental design, e.g. *crossing the ten.*

3.2 Participants

3.2.1 Groups

In designing an experiment where comparisons have to be made between dyslexics and non-dyslexics, controls matched by chronological age are included among the participants. If a difference in performance is a distinctive characteristic of dyslexics, they will also differ from controls matched by spelling age (SA controls), indicating some kind of deficit as opposed to a delay. Two control groups were therefore required.

Conventionally dyslexics are compared with chronological age controls (CA controls) and reading age controls (RA controls). The reason for this is to provide direction of causality. If dyslexics and CA controls are on a level there is no problem. But if dyslexics perform worse on some task, is this due to their poor reading? If so RA controls will be at the same level. Nicolson and Fawcett (1995) have found tasks where dyslexics scored lower than RA controls.

Many who used this methodology equated dyslexia with poor reading, but poor spelling is a much better indication. It is well known within dyslexia teaching that progress in reading skills often outstrips that of spelling, which is the more resistant to improvement after specialised teaching input. In using SA controls the intention was to look for evidence of a deficit as opposed to delay – in which case SA controls would perform better than dyslexics. To our surprise they didn't (apart from the use of the algorithm hypothesis – see Chapters 6 and 10) so it was decided to leave out SA controls for the other three operations. With regard to the direction of causality problem, if the poorer results for dyslexics were due to their slowness in reading, they would be slower in search time – so we had to build in a search time procedure for the other three operations which also measured speed of motor responses.

A spelling test was chosen rather than a reading test because it is widely regarded that children can have dyslexic-type problems and still be adequate readers (Miles et al., 1998).

It is preferable to select controls matched on spelling age rather than reading age. Although many researchers have equated dyslexia with 'poor reading', this does not do justice to the complexity of the interaction between biological abnormality, cognitive deficit and behavioural manifestations (Frith, 1995), nor to the many weaknesses other than poor reading which have been reported by Nicolson and Fawcett (1995). In addition, if dyslexics are selected only on the basis of poor reading there is a serious risk of false negatives (children who show many of the typical dyslexic signs but who have learned to read adequately). This risk is particularly large if, as in the present research, the participant population is drawn from schools that specialise in the teaching of dyslexic pupils. In contrast, in view of the indisputable difficulties experienced by dyslexics in learning to spell, any child within the age-range of the present research who turned out to be an adequate speller could for that very reason be at most only mildly dyslexic.

In Stage 2, the findings from the SA controls enabled a decision to be made to exclude them from the Division, Addition and Subtraction experiments and focus on a comparison of the dyslexic group with the CA control group (referred to as the control group in Stage 2). Another refinement was the addition of a Reading/Key Search

Reaction Time Test, included within the design for Stage 2 so as to form an experimental safeguard. This test enabled a study of whether the performance of the dyslexic participants was affected by a slower reading rate and slower motor skill level. This design enabled the experimenter to arrive at a decision as to whether or not to adapt the findings for Stage 2.

In order to minimise variables that could affect the results, the following were controlled for in both stages: age, sex, geographical location and socio-economic status. The age of participants is detailed in sections 3.3.1 (Stage 1) and 3.4.1 (Stage 2). Sex is controlled for in that the participants were only boys. The controls were from the same geographical locations (rural Southern England including schools in Somerset, Wiltshire, Berkshire and Hertfordshire). Participants were chosen from private schools to control for socio-economic status. (See section 3.2.6 for further details on the schools used.)

Stage 1

Participants were selected in such a way that they could be assigned to one of three groups, namely:

- Dyslexic – composed of dyslexic participants.
- CA controls – control participants matched with the dyslexic group for age and reasoning ability.
- SA controls – a second control group matched with the dyslexic group on spelling age and reasoning ability.

From a large pool of available participants who had been given the initial tests outlined in section 3.2.3, 90 participants were selected from 3 different schools, some of which took only dyslexic children and others that took dyslexics and other children. Each experimental group contained 30 participants that satisfied the criteria specified in section 3.2.4.

Stage 2

Participants were selected in such a way that they could be assigned to one of two groups, namely: dyslexics and controls matched in pairs for age and intelligence.

From a large pool of available participants who had been given the initial tests outlined in section 3.2.3, 60 new participants were selected from 7 schools, some of which took only dyslexic children and others that took dyslexics and other children. Each experimental group contained 30 participants that satisfied the criteria specified in section 3.2.4.

3.2.2 Age-bands for the Four Operations

Participants in the dyslexic and CA control groups were identified in three age bands as presented in Table 3.1. Participants from the SA matched control group are not represented here because of the very nature of their being matched with the dyslexics by spelling age which did not necessarily correspond with the age bands presented in Table 3.1.

Table 3.1 Chronological age range within each age band

Age Band	*Age (years and months)*
Young	9:5 – 11:4
Medium-age	11:5 – 13:4
Old	13:5 – 15:4

Throughout the research reference will be made to the three age bands: young, medium-age or medium age band and old based on the ages given in Table 3.1. In some places these terms are abbreviated to 'Y', 'M' and 'O'.

It was decided not to include any dyslexic participant under 9:5 years in the young age band because they would be most likely matched with a participant of

approximately 7 years 1 month on spelling age. The SA controls would therefore have their chronological ages around their spelling ages, in line with the criteria of this research. Multiplication and division number facts may not be taught to children who are younger than 7 years old.

Each age band represents two years, spanning both the Junior and Senior stages of schooling. At the ages of 15/16 years students take important examinations (GCSE) to help determine their future career paths. This research was designed to examine the responses of students who were already participating in their GCSE coursework. Comparisons could then be made within age bands and also across them in the final analysis of results. By this means, the performance of the dyslexic participants could be accurately observed over a six-year age period.

Different participants were used in each age band. This was not a longitudinal study.

The control participants are sometimes referred to as non-dyslexics. Both terms are used interchangeably.

3.2.3 Testing

Preliminary tests were conducted in order to identify suitable participants for the research. All prospective participants undertook the following tests:

The Schonell S1 spelling test (Schonell and Schonell, 1952)
This test was straightforward to administer and was given to a large group simultaneously. The participant score on this test (number of words spelled correctly) was used to match participants across the experimental groups.

The Raven Standard Progressive Matrices (SPM) (Raven, 1958)
All the participants chosen were tested on reasoning power in order to forestall any possible objection that differences in the final results were due to this. In order to control for reasoning power Raven Matrices was chosen. In the case of those aged less than 8 years in the SA control group, the Raven Coloured Matrices (Raven, 1962)

was used. Total raw scores were used to match participants across the experimental groups.

The Bangor Dyslexia Test (BDT) (Miles, 1982, 1997)

A test was required that could distinguish dyslexics from non-dyslexics. The BDT was chosen for two reasons: firstly for its scientific rigour, and secondly its clinical background.

> A recognisable pattern of difficulties was clearly emerging: in particular the spelling of many subjects seemed bizarre; many of them continued to confuse 'b' and 'd' at a relatively advanced age; and they were regularly showing some or all of the difficulties ... – problems over left-right, over repeating polysyllables, over subtraction and tables, and over repeating the months of the year and strings of digits. I was also confronted with clear evidence that the condition sometimes runs in families. Confirmation of these ideas came not only from discussions with colleagues and a study of the relevant literature, but from the fact that what I said was clearly making sense both to the children and to their parents. (Miles, 1993, p. 16)

In addition to any previous independent diagnosis, the researcher assessed each participant for any possible indication of dyslexia. The BDT was chosen in order make sure that none of the potential controls showed typical dyslexic behaviour, thus making sure that no dyslexic was included in the control group or vice versa.

When scoring the test, two sub-items, Subtraction and Tables, were excluded. This was in order to prevent inclusion of participants who had been selected because they were, in part, weak at the two sub-items that involved mathematics.

3.2.4 Criteria for Selection for the Four Operations

The following criteria were specified in the selection of participants.

No participant was included if they had:

(1) a gross physical handicap;

(2) major problems of adjustment;

(3) participants had to have a score on the Raven Standard Progressive Matrices (Raven, 1958) or on the Raven Coloured Matrices (Raven, 1962) that was above the 50th percentile for his age except in the case of one participant in the old CA control group whose score placed him on the 40th percentile. This was to ensure that the results were not affected by extraneous factors such as low reasoning ability.

Criteria for acceptance into the *dyslexic* group:

(1) a spelling age on the S1 spelling test (Schonell and Schonell, 1952) of not less than 18 months below chronological age, or 2 years below chronological age for participants over 13 years of age;
(2) at least 3.5 positive indicators out of 8 on the Bangor Dyslexia Test (Miles, 1982, 1997).

Criteria for acceptance into the *control* groups:

(1) a spelling age equal to or above chronological age or within 6 months of it, or in the case of children in the old age band, a spelling age of at least 12 years;
(2) not more than 2.5 positive indicators out of 8 on the Bangor Dyslexia Test.

To prevent any possible effects of gender difference it was decided to select only male participants.

3.2.5 Justification for Boundaries

The choice of the positive indicators on the Bangor Dyslexia Test was adjusted in line with this research concerning mathematics. Normally a score of 4 or more pluses on the Bangor Dyslexia Test gives an indication of dyslexia (Miles, 1982, 1997). In the Bangor Dyslexia Test there are two sub-tests connected to mathematics – Subtraction and Tables. If these two tests had been included in the selection process this could have attracted the legitimate criticism that a participant may have been chosen because of their poor scoring on the mathematical items of this test and thus the research results would have been biased by the selection procedure for participants. It

was therefore decided not to include the results for 'Subtraction' and 'Tables' within the Bangor Dyslexia Test, leaving 8 instead of 10 sub-tests for scoring. A buffer zone on the Bangor Dyslexia Test was created so that there was a difference of at least one positive indicator between the dyslexic and non-dyslexic participants.

Setting the boundaries for spelling

Two participants in the young and medium-age dyslexic groups scored 18 months behind their chronological age on the spelling test. Both scored 51 and 55 respectively on the SPM and they gained at least 3.5 pluses on the BDT where both mathematical tests had been excluded. It was decided to set the spelling boundary for the young and medium-age dyslexics at 18 months behind chronological age as a result.

One participant in the old age band scored 72 on the spelling test making his spelling age 2 years behind his chronological age. He scored 56 on the SPM and on clinical grounds was obviously dyslexic, scoring at least 3.5 pluses on the BDT where the two mathematical items had been excluded. On this basis it was decided to set the spelling boundary at a maximum of 72 and at two years below chronological age for the old participants in the dyslexic group.

The gap between CA (mean chronological age of the participant during initial testing for selection as a suitable candidate) and *CA* (the mean chronological age during the mathematical testing) was sometimes as much as 20 weeks for practical reasons. It was therefore decided that, for valid comparisons to be made, the ages at which the calculation items were given should be the ones used. Thus it was considered appropriate to use the mean of the participant's age over the mathematical testing period.

The CA controls were matched with specific dyslexic partners on age, at time of mathematical testing, and intelligence (raw score). At most the CA control participant was no more than 2 months older than the dyslexic partner. At most the dyslexic was no more than 4 months older than his control partner. The overall mean age for the dyslexics in a group/age band was higher than the mean age for the CA controls.

In Stage 2 there were no cases where the CA controls were ahead of the dyslexics by more than 2 points on the SPM raw score. The overall mean raw scores were higher for the dyslexics in all age bands than the control groups.

Each participant thus experienced many preliminary series of tests before he was chosen for the main test in which his responses and answers would be recorded. Individual participant details for Stage 1 and Stage 2 are given in Appendix A (Tables A.1 to A.15)

3.2.6 Schools

All participants were chosen from seven private schools, two of which accepted only dyslexics and five of which had an established dyslexia unit. It was decided that these schools were sufficiently alike in socio-economic status and methods of teaching to justify the research selection criteria. Moreover, in the case of the dyslexic participants, their poor spelling performance could not be due to a lack of opportunity to learn.

This research was conducted in two stages with two different groups of participants. Stage 1 involved a study of performance on Multiplication and Stage 2 was concerned with experimentation on Division, Addition and Subtraction, thus rounding off a study of the four mathematical operations.

3.3 Stage 1 – The Multiplication Experiment

3.3.1 Participant Details

The mean chronological ages and spelling ages for the 9 groups are given in Table 3.2 and Table 3.3. Chronological ages and spelling ages are given in years and months, with standard deviations in months.

Table 3.2 Mean chronological ages, spelling ages, Raven Matrices quotient and Bangor Dyslexia Test pluses for the 3 groups and 3 age bands in Stage 1.

Group	*Type*	*Age Band*					
		Young		*Medium*		*Old*	
		Age or Matrices Score	*Standard deviation*	*Age or Matrices Score*	*Standard deviation*	*Age or Matrices Score*	*Standard deviation*
Dyslexics	CA	10:7	+/– 3.65	12:7	+/– 8.85	14:3	+/– 6.34
	SA	7:10	+/– 5.17	9:11	+/– 8.60	11:4	+/– 9.73
	MS	114	+/– 4.95	114	+/– 8.71	114	+/– 8.03
	BDT	7.3	+/– 1.58	5.8	+/– 1.40	5.8	+/– 1.44
	BDT*	5.75	+/– 1.34	4.85	+/– 1.08	4.7	+/– 1.16
CA controls	CA	10:0	+/– 4.24	12:1	+/– 8.11	14:1	+/– 5.52
	SA	10:11	+/– 7.49	12:11	+/– 5.55	13:11	+/– 4.23
	MS	113	+/– 6.08	112	+/– 7.62	110	+/– 4.32
	BDT	2.45	+/– 0.80	2.75	+/– 0.35	1.85	+/– 0.85
	BDT*	1.6	+/– 0.61	2.05	+/– 0.37	1.4	+/– 0.77
SA controls	CA	7:8	+/– 5.61	8:9	+/– 8.16	9:5	+/– 8.33
	SA	7:10	+/– 4.87	9:10	+/– 7.66	11:4	+/–10.23
	MS	111	+/–10.35	120	+/– 6.64	122	+/– 6.30
	BDT	2.75	+/– 0.82	1.7	+/– 1.09	1.75	+/– 1.03
	BDT*	1.6	+/– 0.97	1.1	+/– 0.77	1.1	+/– 0.91

Note.

CA = chronological age (years and months). Standard deviation is given in months.

SA = spelling age (years and months). Standard deviation is given in months.

MS = performance on Raven Standard Progressive Matrices or Raven Coloured Matrices.

BDT = Bangor Dyslexia Test.

BDT* = Bangor Dyslexia Test with Subtraction and Tables sub-items omitted.

Thus the dyslexics were marginally older than the CA controls and had marginally higher scores on the Raven Matrices. Any respect in which they performed worse than

the CA controls could not therefore be the result of differences in age or reasoning powers.

Table 3.3 Mean chronological ages, spelling score, Raven Matrices score and Bangor Dyslexia Test pluses for the 3 groups and 3 age bands in Stage 1

Group	*Type*	*Age Band*					
		Young		*Medium*		*Old*	
		Age or Score	*Standard deviation*	*Age or Score*	*Standard deviation*	*Age or Score*	*Standard deviation*
Dyslexics	CA	127.3	+/– 3.65	151.2	+/– 8.85	171.3	+/– 6.34
	SS	27.6	+/– 5.17	48.8	+/– 8.60	63.3	+/– 9.73
	MS	42.3	+/– 3.16	47.0	+/– 4.03	49.6	+/– 3.47
	BDT	7.3	+/– 1.58	5.8	+/– 1.40	5.8	+/– 1.44
	BDT*	5.75	+/– 1.34	4.85	+/– 1.08	4.7	+/– 1.16
CA controls	CA	119.7	+/– 4.24	144.6	+/– 8.11	169.3	+/– 5.52
	SS	59.1	+/– 7.49	79.2	+/– 5.55	89.1	+/– 4.23
	MS	38.3	+/– 4.69	44.6	+/– 2.55	49.3	+/– 4.08
	BDT	2.45	+/– 0.80	2.75	+/– 0.35	1.85	+/– 0.85
	BDT*	1.6	+/– 0.61	2.05	+/– 0.37	1.4	+/– 0.77
SA controls	CA	91.8	+/– 5.61	105.2	+/– 8.16	112.5	+/– 8.33
	SS	27.8	+/– 4.87	48.2	+/– 7.66	63.2	+/–10.23
	MS	25.9	+/– 5.65	38.2	+/– 7.81	42.4	+/– 6.57
	BDT	2.75	+/– 0.82	1.7	+/– 1.09	1.75	+/– 1.03
	BDT*	1.6	+/– 0.97	1.1	+/– 0.77	1.1	+/– 0.91

Note.

CA = chronological age (months) with standard deviation in months.

SS = spelling score.

MS = raw score on the Raven Standard Progressive Matrices or Raven Coloured Matrices.

BDT = Bangor Dyslexia Test.

BDT* = Bangor Dyslexia Test with Subtraction and Tables sub-items omitted.

Up to a maximum of eight weeks elapsed between the time a participant was first asked to take part in a group administration of the Schonell S1 spelling test (Schonell and Schonell, 1952), through to the Raven Standard Progressive Matrices (1958) or Raven Coloured Matrices (1962) and individual observation in the Bangor Dyslexia Test (Miles, 1982, 1997).

3.4 Stage 2 – Division, Addition and Subtraction Experiments

3.4.1 Introduction and Participant Details

The Multiplication experiment gave rise to findings on correctness and speed of response. The results are discussed in Chapters 6 and 10. Further questions arose, regarding the dyslexics' performance on the other three operations in mathematics. Would the trend in the results of the Multiplication experiment be evident in addition, subtraction and/or division? See Chapter 16 for a comparison of the four operations.

Stage 2 preparations began after testing for Stage 1 was completed and the results analysed. Computer software was programmed with a similar format in mind but with a few modifications, which are outlined in this chapter.

Based on the findings for Stage 1 – Multiplication, the SA controls were not included in Stage 2. The overall results for the SA controls had indicated that their performance was both slower and less accurate than the dyslexic group in all age bands.

In Stage 2, division was to be tested and it was clear that in the school curriculum the subject of division would not be adequately covered until most pupils were in their ninth year of age. This linked in with the young age band in this research but by the very nature of matching on spelling age it was found that several participants would need to be sought who fell below this age in all the three potential spelling age groups. This was due to the low spelling scores of dyslexics in all age bands. An example of this can be seen in the old dyslexic group where one 15-year-old gained a spelling score which placed his spelling age in the six-year age bracket. To find and test a spelling age matched participant in his sixth year was not appropriate. On this

basis it was decided to include CA controls and dyslexics in Stage 2 and exclude SA controls.

It was decided to question participants on one particular operation at a time instead of presenting a mixture of types of question (on division, addition and subtraction within the one experiment). Dyslexics are well known for confusing mathematical symbols.

Stage 2 group mean chronological ages, spelling ages and Raven Matrices quotients are given for each of the age bands in Table 3.4. Additionally, spelling scores and Raven Matrices scores are given in Table 3.5.

Table 3.4 Mean chronological ages, spelling ages, Raven Matrices quotient and Bangor Dyslexia Test pluses for the 2 groups and 3 age bands in Stage 2

Group	*Type*	*Age Band*					
		Young		*Medium*		*Old*	
		Age or IQ	*Standard deviation*	*Age or IQ*	*Standard deviation*	*Age or IQ*	*Standard deviation*
Dyslexics	CA	10:7	+/–7.80	12:3	+/– 4.50	14:9	+/– 4.31
	CA	10:8	+/– 7.47	12:5	+/– 5.46	14:10	+/–4.38
	SA	7:10	+/– 9.42	9:0	+/–10.54	10:7	+/–15.90
	MS	115	+/– 6.35	119	+/– 4.67	117	+/– 8.28
	BDT	6.95	+/– 1.59	6.3	+/– 1.46	5.75	+/– 0.98
	BDT*	5.15	+/– 1.38	4.95	+/– 1.09	4.55	+/– 0.76
Controls	CA	10:5	+/– 7.29	11:11	+/– 5.83	14:8	+/– 3.71
	CA	10:7	+/– 8.01	12:2	+/– 5.73	14:9	+/– 3.97
	SA	11:8	+/– 9.29	12:11	+/– 8.59	13:10	+/– 3.78
	MS	112	+/– 7.54	111	+/– 4.33	110	+/– 8.79
	BDT	2.05	+/– 0.90	1.8	+/– 0.86	2.45	+/– 0.50
	BDT*	1.45	+/– 0.86	1.1	+/– 0.61	1.9	+/– 0.66

Note.

CA = chronological age (years and months).

CA = mean age of the participants when performing the mathematical experiments.

SA = spelling age (years and months).

MS = performance quotient on Raven Standard Progressive Matrices or Raven Coloured Matrices.

BDT = Bangor Dyslexia Test.

BDT* = Bangor Dyslexia Test with Subtraction and Tables sub-items omitted.

In the case of CA, *CA* and SA the standard deviation is given in months.

In each age band the dyslexic group chronological mean age is therefore higher than the controls and their mean scores are higher on the test of reasoning power.

Table 3.5 Mean chronological ages, spelling scores, Raven Matrices scores and Bangor Dyslexia Test pluses for the 2 groups and 3 age bands in Stage 2.

Group	*Type*	*Age Band*					
		Young		*Medium*		*Old*	
		Age or Score	*Standard deviation*	*Age or Score*	*Standard deviation*	*Age or Score*	*Standard deviation*
Dyslexics	CA	126.7	+/– 7.80	147.3	+/– 4.50	177.1	+/– 4.31
	CA	128.3	+/– 7.47	148.6	+/– 5.46	178.1	+/– 4.38
	SS	28.4	+/– 9.42	40.4	+/–10.54	56.0	+/–15.90
	MS	42.4	+/– 6.04	49.4	+/– 2.55	51.2	+/– 3.88
	BDT*	5.15	+/– 1.38	4.95	+/– 1.09	4.55	+/– 0.76
Controls	CA	124.7	+/– 7.29	143.3	+/– 5.83	176.2	+/– 3.71
	CA	127.1	+/– 8.01	146	+/– 5.73	177.3	+/– 3.97
	SS	66.9	+/– 9.29	79.4	+/– 8.59	87.6	+/– 3.78
	MS	39.7	+/– 4.32	44.5	+/– 2.55	48.1	+/– 4.07
	BDT*	1.45	+/ 0.86	1.1	+/– 0.61	1.9	+/– 0.66

Note.

CA = chronological age (months).

CA = mean age of the participants when performing the mathematical experiments.

SS = spelling score.

MS = raw score on the Raven Standard Progressive Matrices or Raven Coloured Matrices.
BDT* = Bangor Dyslexia Test with Subtraction and Tables sub-items omitted.
In the case of CA and *CA* the standard deviation is given in months.

3.5 Apparatus

The apparatus section of this chapter is subdivided into the two main experimental stages presented in 3.6 and 3.7 and the apparatus used is detailed.

3.6 Stage 1 – Apparatus for the Multiplication Experiment

3.6.1 Introduction

After selection in accordance with the criteria outlined above, each participant was given instruction in the use of an Amstrad PPC 640D portable computer with a white on black Sinclair S-12 MM screen placed centrally in front of him. The screen was positioned so that any glare from lighting or sunlight was minimised. The Amstrad keyboard was so placed as to position the participant centrally to the number board located on the right of the keyboard.

3.6.2 Software Design

The following disks were used:

(1) Microsoft MSDOS operating system
(2) program working disk
(3) storage disks
(4) Lotus 123
(5) WordStar Express
(6) SPSS Software Package

3.6.3 The Program Working Disk

This disk was created with the following format:

(1) A set presentation of 10 multiplication questions presented centrally, one at a time on the monitor, in the following sequence:

$1 \times 1, 1 \times 2, 1 \times 3, 1 \times 4, 1 \times 5, 1 \times 6, 1 \times 7, 1 \times 8, 1 \times 9, 1 \times 10$

The answers indicated by the participant would enable him to become familiar with the position and operation of all the keys.

(2) 10 random (pre-set) multiplication questions presented individually on the screen as a preliminary practice in the following sequence:

$3 \times 4, 5 \times 10, 6 \times 2, 1 \times 12, 5 \times 3, 7 \times 5, 10 \times 10, 8 \times 4, 11 \times 9, 12 \times 2$

(3) The main test of 144 random (pre-set) multiplication questions presented individually on the screen. Tables of Random Permutation (Moses and Oakford, 1963) were used to select at random the order of the questions, using the 144-squared multiplication grid as a preparation pad.

As an example, the number 76 is given as the first number on page 141 of Table 7 (Moses and Oakford, 1963), so this was transposed to box 76 on a pre-numbered multiplication grid. The grid was so numbered as to start with 1×1 as box 1, following on in sequence horizontally to 12×1 as box 12. The maximum number of boxes was 144. Where Table 7, followed sequentially, showed a number greater than 144, this was discarded. Because box 76 was the first to be selected from the random numbers presented in Table 7 this was found to correspond to 4×7 (taking the horizontal reading first) and hence this was to become the first question for all participants in Trial 1 as part of the main experiment. A second trial was prepared whereby all the questions were rotated around a central point so that the first question 4×7 in Trial 1 became

the last question; the second question in Trial 1 became the second to last question in Trial 2 and so on.

(4) A maximum presentation time of 22 seconds – from the pilot study it was established that 99.8% of responses, whether accurate or inaccurate, were made within the range of 0–20 seconds. Accordingly, it was decided to adopt a 20-second upper limit with a 10% tolerance margin. This produced a time of 22 seconds as the maximum permitted time for a response by any of the participants in the research.

The alternative was to adopt an arbitrary time limit, determined a priori, such as 60 seconds. This would have created a potentially unmanageable set of testing procedures that could have lasted for over two hours for any participant in the worst-case scenario.

The data produced, it is contended, reveals a very small percentage error. This error would be a Type II error in that it would create false negatives that would reduce the probability of finding significant differences or factors in the research.

This conservative approach has been adopted throughout the research and, as such, may explain a small number of cases where a significant difference has failed to be substantiated. This was felt to be the less damaging alternative of the two options, as it is empirically based and, if anything, biases the research towards the null hypothesis.

The disk was created using Turbo Basic, with the help of a computer programmer/analyst. The programmer analysed the program development to achieve the objectives. The algorithm was written and subsequently coded, debugged and modified as requirements were refined.

3.7 Stage 2 – Apparatus for the Division, Addition and Subtraction Experiments

3.7.1 The Computer and Base

Each participant was given instruction in the use of an AST Bravo NB 4/25s Laptop Computer with an attached numeric keypad. The numeric keypad became integral with the computer by placing it on a specially designed base that centrally located the device to enable quick assembly and consistent location close to and in front of the screen. The base was placed over the Laptop keyboard to prevent the participants from selecting irrelevant keys. The base was padded to help with comfort of prolonged use.

3.7.2 The Numeric Keypad

This pad comprised the numbers 0 through to 9 with a large "Enter" key laid out in the regular figuration presented on the right-hand side of a full computer keyboard. Other keys on the keypad, which were unnecessary for this research, were covered over to prevent visual distraction and mistaken use. Additional precautions were taken by disconnecting the "Delete" key, as it lay very close to the "Enter" key – it was felt necessary to do this because under timed conditions a participant wishing to operate quickly may inadvertently 'catch' this key.

3.7.3 Software Design

The following disks were used:

(1) WordPerfect word processor package.

(2) React and Program disks, written in Turbo Basic (aided by X Tree Pro) by a computer programmer, were loaded on the hard drive for immediate boot up.

(3) Storage of data was automatic onto the hard disk under the heading 'Data Storage'.

(4) SPSS statistical analysis program.

(5) Presentation of statistical data on Microsoft Word and Microsoft Excel.

3.7.4 Program Working Disc

This was designed in the same way as for Stage 1. Pre-tests were included for each operation in Stage 2; these were the set pre-test and then a random pre-test, like Stage 1. The main test consisted of 150 questions (Addition and Subtraction experiment) and 144 questions for the Division experiment, presented individually on the screen. The order of the questions was decided in the same way as for Stage 1 – Tables of Random Permutation (Moses and Oakford, 1963) were used to select at random, the order of the questions.

3.7.5 Program Working Disk – Improvements

As a result of the practical experience gained from Stage 1, improvements were made to the new program for ease of use and analytical measurement. The improvements comprised:

(1) An auditory single 'beep' to signal the appearance of the next question. This was included for the purpose of focusing the participant's mind on the start of the next question.

(2) The screen presentation was changed from black to white to alert the participant to the imminent change of question. This coincided with the beep. Thus a multi-sensory visual and auditory primer was given to participants on each question.

(3) It was decided to enlarge the number screen presentation to a digit height of 1.5 cm for greater clarity. The digits were positioned centrally from left to right but 4 cm down from the top of the screen and 9 cm up from the bottom of the screen in order to cater for the Base and Numeric Keypad location. The screen width was 19.25 cm and the screen height was 14.5 cm. The digits were therefore presented (taking their base) 5.5 cm from the top of the screen. There was a 0.5 cm gap between digits where 2 digits were included in a number.

(4) No changes were made to the interval between question presentations, the maximum 22 second presentation or the halfway point in the main experiment.

(5) Measurement considerations: as a result of using a more powerful computer, greater analysis of data was possible. Each time the participant depressed a key, the computer timed this response, thus enabling measurements between:

(a) screen question display and first key; first key and "Enter" response;

(b) screen question display and first key; between first key and second key; between second key and "Enter" response;

(c) screen question display and first key; between first key and second key; between second key and third key; between third key and "Enter" response where applicable;

(d) the total response time between screen question display and "Enter";

(e) screen question display to no response within 22 seconds (called a 'miss');

(f) screen question display to "Enter" where not enough keys were pressed for the correct answer (called a 'pass').

The next four sections (3.8 to 3.11) detail the methodology for each of the four operations.

3.8 The Multiplication Experiment

3.8.1 Instructions for Program Working Disk (Main Instructions)

With the participant comfortably positioned in front of the computer, the experimenter keyed in the preliminary details of name and age in years and months for accurate identification later. The participant was instructed as follows:

(1) 'You will be working on the computer which will be timing how quickly you respond.'

(2) 'The computer will ask you questions on the screen here [experimenter indicates the centre of the screen] and you then press the answer in here [experimenter indicates the numeric keypad]. Once you have pressed your answer, you must press the "Enter" key immediately because you are still being timed. It is called 'enter' because you want to enter your answer into the computer's memory.'

(3) 'The clock inside the computer will start timing you from the moment the question appears on the screen. It stops timing you as soon as you press the "Enter" key.'

(4) 'Think of the answer as quickly as possible and always have your hand or hands at the ready.'

(5) 'If you really don't know the answer, press the "Enter" key alone. That question will disappear and the computer will ask you the next question, so be ready for it. We would like you to answer as many of the questions as possible for this research.'

(6) 'After a reasonable length of time, if you have not answered a question, it will disappear – so keep your eyes on the screen so that you don't miss the next question.'

At this point the experimenter sequentially showed the participant each number on the pad, naming each one as it was pointed out.

(7) 'The "Delete" key does not work. If you make a mistake call out your answer and I will make a note of it.'

(8) 'When I say Ready, Steady, Go, press the "Enter" key to start.'

The experimenter now exchanged places with the participant, who then took his seat in front of the screen and numeric keypad, and the testing began.

For all four operations the question remained on the screen, while the participant typed in his response (this was shown on the computer screen), until the participant pressed the "Enter" key or "Missed" if 22 seconds had elapsed.

3.8.2 Set Pre-test

The participant was asked the ten questions in ascending order so that he could familiarise himself with the positions of the numbers on the keys. This pre-test also allowed the participant time to consider which hand or hands he was going to use before the main test began. The participant could use this opportunity to develop an

understanding of the way in which he needed to press the keys so that his answers would appear on the screen.

The 10 multiplication questions were presented one at a time on the monitor, in the following sequence:

$$1 \times 1, 1 \times 2, 1 \times 3, 1 \times 4, 1 \times 5, 1 \times 6, 1 \times 7, 1 \times 8, 1 \times 9, 1 \times 10$$

The answers indicated by the participant would enable him to become familiar with the position and operation of all the keys.

On completion of this pre-test the following instruction was displayed on the screen:

> Question ONE will start when you press "ENTER".

3.8.3 Random Pre-test

This pre-test enabled the participant to become familiar with the kind of variation that he was about to encounter in the main experiment. These ten questions were specifically asked because they included all the numbers from 1 to 12.

The 10 random (pre-set) multiplication questions were presented individually on the screen as a preliminary practice in the following sequence:

$$3 \times 4, 5 \times 10, 6 \times 2, 1 \times 12, 5 \times 3, 7 \times 5, 10 \times 10, 8 \times 4, 11 \times 9, 12 \times 2$$

On completion of this pre-test the following instructions were displayed on the screen:

> You have completed the short section [This was visible for only one second before the next instruction was given below.]
>
> Question ONE will start when you press "ENTER".

3.8.4 Main Experiment – Multiplication

The main test of 144 random (pre-set) multiplication questions were presented individually on the screen. See section 3.6.3 for details on how the order of questions was chosen.

3.8.5 Question Order for Multiplication

The questions and the order in which they were presented are in Appendix B (Table B.1).

The first 72 questions were displayed on the screen, one at a time, and the participant responded in accordance with his instructions. This was followed by an interval of a few minutes. The interval was followed by questions 73 to 144. On reaching question 144, the participant had completed Trial 1.

Trial 2 then took place during the following two weeks. The same procedure, including the two pre-tests, was carried out to familiarise the participant with the task in hand. This time in the Main Experiment, the questions were pivoted around a central point, as explained in section 3.6.3.

The researcher carried out all the testing for this research, seated either to the side of the participant some four feet away, or slightly behind the participant at the same distance. This was in order to establish a situation in which the participant would not be too self-conscious or even be aware that he was under observation, which could have had an adverse effect upon his concentration and responses. It was also necessary for the researcher to see the screen. She therefore sat in an unobtrusive position, but within sight of the computer.

At the end of each testing session the participants were thanked for their help but no comments were made on their performance.

3.8.6 Recording Components by the Computer

The computer had been programmed to store the following data:

(1) The question asked, the correct answer, the answer given and the time taken for each individual key to be pressed in response, as well as a total response time to the nearest one hundredth of a second.

(2) A total breakdown of the four different types of response, namely right answers, wrong answers, passed questions and missed questions for the two pre-tests and the two halves of the Main Experiment.

All the data were stored in ASCII format, which was 'comma delimited' as a preparation for analysis by the SPSS software package.

3.8.7 Response Categories for Analysis

(1) *Correct* – where a participant answered correctly within the 22-second time-limit available the response was counted as 'correct'.

(2) *Incorrect* – where a participant responded within the 22-second time-limit available but typed in the wrong answer prior to pressing "Enter".

(3) *Missed* – where a participant did not press "Enter" before the 22-second time-limit. The participant may or may not have supplied an answer.

(4) *Passed* – where a participant decided that they were unable to answer a question, they were able to press "Enter" and move directly on to the next question. The time taken to make this decision was recorded.

Results for this experiment are given in Chapter 6 (Correctness) and Chapter 10 (Speed). The researcher explored the number of individual observations where young dyslexics in particular realised that they didn't have any prospect of giving an answer to a multiplication question. They tended to pass relatively quickly after stimulus presentation whereas for all other groups this tendency was not so pronounced. Accordingly two measures of central location were used – a median value for each of the group age bands and a compensated mean time, where all inaccurate responses were coded at the maximum time out level of 22 seconds.

Where a particular multiplication table has been referred to in the results, a multiplication sign has been placed after the table number, e.g. 2×, meaning in this case the two times multiplication table.

3.9 The Division Experiment

As with the Multiplication experiment, Division shared many similarities, namely:

(1) The use of the same combinations of numbers ranging from 1 ÷ 1 through to 144 ÷ 12. All possible division questions between this range were used.
(2) Only division questions resulting in whole number answers with no remainders were used.
(3) The method of random selection for the order of the questions was the same.
(4) The same number of questions (144) with an interval at the halfway point was used.

The concept of division shares a similar foundation to that found in multiplication. The two operations are interchangeable, for example, $5 \times 7 = 35$ and $35 \div 7 = 5$.

3.9.1 Instructions for Program Working Disk (Main Instructions)

The experimenter keyed in the preliminary details of participant name, age and school, for accurate identification later. In addition the experimenter made a note of these details in a separate notebook together with participant's date of birth and date of testing. Each participant had a page entry ready for observations.

The experimenter then explained the details of the test situation and gave the following instructions to the participant:

(1) 'You will be working on the computer that will be timing how quickly you respond to questions on division.'
(2) 'The computer will ask you a question on the screen here [experimenter indicates the centre of the screen where the characters will appear] and you then press the answer in here [experimenter indicates the numeric keypad].'

(3) 'Once you have pressed in your answer, you must then press the "Enter" key [experimenter points this out] immediately to stop the clock. You are being timed from the moment the question appears on the screen to the time when you type in your answer and press the "Enter" key.'

(4) 'If you really don't know the answer and can't work it out, you may press the "Enter" key by itself to make that question go away. The next question will then come up. This will be signalled by a beep to let you know it is there.'

(5) 'There is a maximum time period of about 20 seconds to answer a question. If you take longer than this, the question will go away and the next question will appear.'

At this point the experimenter sequentially showed the participant each number on the numeric keypad from 0 through to 9, naming each one as it was pointed out. The "Enter" key was clearly indicated and named again.

(6) 'The "Delete" key has been instructed not to work. If you make a mistake, hard luck, there is nothing we can do. You then move on to the next question by pressing the "Enter" key.'

(7) 'Like a football match there are two short warm up sessions, and then two halves with an interval in the middle. We will take a short break when you get to the interval.'

(8) 'Remember that you are being timed so think of the answer as quickly as you can.'

The participant then had an opportunity to clarify any part of the instructions and then changed places with the experimenter in order to begin. The experiment was initiated by pressing the "Enter" key once.

3.9.2. Set Pre-test

The Set pre-test was designed to encourage sequentially increasing responses from the participant so that he was able to become familiar with the type of division questions and experience pressing different numbers on the numeric keypad. Nine questions

were prepared designed to evoke responses ranging from 1 through to 9 on the numeric keypad.

The question order was:

$$1 \div 1, 2 \div 1, 3 \div 1, 4 \div 1, 5 \div 1, 6 \div 1, 7 \div 1, 8 \div 1, 9 \div 1$$

On completion of this pre-test the following instruction was displayed on the screen:

> Question ONE will start when you press "ENTER".

3.9.3 Random Pre-test

This was a preliminary test of ten questions designed to prepare the participant for the type and range of questions in the main Division experiment. Both questions and answers employed the use of all the division tables (except 8 and 11 due to there being ten questions and 12 division tables) and all of the keys on the numeric keypad (except 8 and 0). These ten questions were repeated at differing points in the Main Experiment.

No question in this experiment involved the use of '0', either in the numerator or denominator, since this would have increased the number of questions beyond a reasonable length. The inclusion of '0' would have added an unwanted variable.

The question order was:

$$8 \div 4, 22 \div 2, 42 \div 7, 5 \div 5, 70 \div 10, 4 \div 1, 81 \div 9, 72 \div 6, 36 \div 12, 15 \div 3$$

On completion of this pre-test the following instructions were displayed on the screen:

> You have completed the short section. [This was visible for only one second before the next instruction was given below.]
>
> Question ONE will start when you press "ENTER".

3.9.4 Main Experiment – Division

The participants needed to press the "Enter" key to signal the start of the Main Experiment. They had been exposed to both pre-tests and had therefore been prepared for the type of questions to come. A total of 144 division questions were presented in random order chosen in the same manner as that for multiplication. After 72 questions the participant reached an interval. On pressing the "Enter" key twice, the second half of the experiment was activated and the remaining questions were displayed one at a time.

Research from Stage 1 showed that the difference in performance from Trial 1 and Trial 2 in the Multiplication experiment was minimal. Statistically the mid-point was taken between the two multiplication trials but was not significant. Due to these results and the fact that each participant was performing on the Addition, Subtraction and Division trials as part of Stage 2, it was felt by the experimenter that each participant had been researched to the appropriate degree. Therefore only one trial was performed for the Division experiment.

3.9.5 Question Order for Division

The questions and the order in which they were presented are in Appendix C (Table C.1).

The participants performed the three experiments (Division, Addition and Subtraction) on different days in order to approach each experiment afresh. The time lapse between experiments ranged from the next day to four weeks.

At the end of each testing session the participants were thanked for their help but no comments were made on their performance.

3.9.6 Recording Components by the Computer

These are the same as for the Multiplication experiment (see section 3.8.6).

3.9.7 Response Categories for Analysis

These are the same as for the Multiplication experiment (see section 3.8.7).

Results for this experiment are given in Chapter 7 (Correctness) and Chapter 11 (Speed).

3.10 The Addition Experiment

3.10.1 Instructions for Program Working Disk (Main Instructions)

The instructions given were the same as those for the Division experiment with the word 'addition' substituted for 'division'.

The participant then had an opportunity to clarify any part of the instructions and then changed places with the experimenter in order to begin. The experiment was initiated by pressing the "Enter" key once.

3.10.2 Set Pre-test

The participant was asked nine questions in ascending order so that he could familiarise himself with the positions of the numbers on the keys. This also allowed the participant the opportunity of finding the most comfortable hand/s to use for responding prior to the Main Experiment. The participant could also become familiar with the presentation and response pattern required.

The addition questions were as follows:

$$1 + 0, 1 + 1, 1 + 2, 1 + 3, 1 + 4, 1 + 5, 1 + 6, 1 + 7, 1 + 8$$

When answering these questions, the participant responded by pressing each number key in sequential order. The nuances of the experimental situation also became apparent to the participant.

On completion of this pre-test the following instruction was displayed on the screen:

> Question ONE will start when you press "ENTER".

3.10.3 Random Pre-test

This pre-test enabled the participant to become familiar with the kind of variation that he was about to encounter in the Main Experiment.

Ten questions were asked in the following order:

$$24 + 2,\ 53 + 8,\ 69 + 4,\ 37 + 26,\ 71 + 7,\ 28 + 6,\ 61 + 8,\ 35 + 57,\ 21 + 3,\ 89 + 2$$

This selection included two questions from each of the five different categories of addition (explained in the Main Experiment), arranged in random order. All of these questions were novel and were not duplicated in the Main Experiment.

On completion of this pre-test the following instructions were displayed on the screen:

> You have completed the short section. [This was visible for only one second before the next instruction was given below.]
>
> Question ONE will start when you press "ENTER".

3.10.4 Main Experiment – Five Categories

The Main Experiment comprised 75 addition questions falling into *five equal categories.* The rationale for the five categories was such that the range of possible addition questions that could have been used was vast. It was decided to study the effect of two main factors: the addition of low compared to high addends, and secondly crossing and not crossing the ten barrier. The ten barrier refers to a change from one group of tens to another, e.g. 19 + 4 = 23, where the original number 19 has one group of ten and the sum of 19 + 4 gives 23 which has two groups of ten. An

addition question such as 23 + 4 does not cross the ten barrier since the resulting sum 27 still retains the same number of tens (namely 2).

Would the size of the addend and/or addition across the ten barrier or within a grouping of ten have a larger effect on dyslexics compared to controls?

Category 1 (for Addition)

This comprised the addition of a low addend (2, 3, 4) to a two-digit number where the sum did not cross the ten barrier. It was labelled LN, an acronym for 'Low addend Not crossing the ten barrier'. Two examples from LN are: 47 + 2 and 64 + 3.

There were 15 questions in this category and each low addend was used five times. The choice of questions in this category was shaped by the constraints that the number system imposed, namely: in the two-digit number the distribution of the 'tens' ranged from the 20s to the 90s and the distribution of the 'units' ranged from 1 to 7.

Category 2 (for Addition)

This category included the addition of high addends (6, 7, 8) to two-digit numbers where the resultant sum did not cross the ten barrier. It was labelled HN, which stands for 'High addend Not crossing the ten barrier'. Two examples from HN are: 92 + 6 and 41 + 8. There were 15 questions in this category and each high addend was used five times.

The choice of questions in this category was shaped by the constraints that the number system imposed, namely: in the two-digit number the distribution of the 'tens' ranged from the 20s to the 90s and the distribution of the 'units' ranged from 1 to 3. The resulting answers had an 8 or 9 in the 'unit' position.

Category 3 (for Addition)

This group contained 15 questions chosen with low addends (2, 3, 4) added to two-digit numbers with the resulting sum crossing the ten barrier. It was labelled LC,

which stands for 'Low addend Crossing the ten barrier'. Two examples from LC are: 29 + 2 and 47 + 4. There were 15 questions in this category and each low addend was used five times.

The choice of questions in this category was shaped by the constraints that the number system imposed, namely: in the two-digit number the distribution of the 'tens' ranged from the 20s to the 80s and the distribution of the 'units' ranged from 7 to 9. The resulting answers had a 1, 2 or 3 in the 'unit' position.

Category 4 (for Addition)

This category included two-digit numbers with high addends (6, 7, 8) and sums which crossed the ten barrier. It was labelled HC, which stands for 'High addend Crossing the ten barrier'. Two examples from HC are: 37 + 6 and 73 + 8. There were 15 questions in this category and each high addend was used five times.

The choice of questions in this category was shaped by the constraints that the number system imposed, namely: in the two-digit number the distribution of the 'tens' ranged from the 20s to the 80s and the distribution of the 'units' ranged from 3 to 9.

Category 5 (for Addition)

This additional category was chosen to represent the hardest questions of all. Their components included the addition of a two-digit number to another two-digit number where the addend units were high (6, 7, 8) and the resulting sums crossed the ten barrier with 'carrying'. It was labelled as the HARD category and two examples are as follows: 65 + 26 and 35 + 28.

The choice of questions in this category was shaped by the constraints that the number system imposed, namely: in the first two-digit number the distribution of the 'tens' ranged from the 20s to the 60s and the distribution of the 'units' ranged from 3 to 9. The addends included numbers ranging from the 20s to the 60s with the distribution of the 'units' ranging from 6 to 8.

Addition questions within each category are given in Appendix D (Table D.1).

3.10.5 Decisions about the Questions

Careful consideration was given to the digits used in the Addition experiment. The experiment did not sample the whole population of addition sums but only a specified selection so as to avoid increasing the variables. A decision was made to restrict the design to certain types of question otherwise the number of sums needed would have been too great. High (large) and low (small) addends were selected as well as sums that crossed and did not cross the ten barrier because they seemed interesting; as a logical consequence, certain other sums were therefore excluded.

Decisions were made to exclude the following:

(1) *Zeros in both the questions and answers.* The ten barrier is the point at which there is a change in the value of the number in the tens position. This point is also indicated by a zero in the units position. Questions were sought which actually involved the 'crossing' of the ten barrier. Additionally a question with a zero in it could, by its very nature, only be included in LN and HC categories, and this would give an unfair advantage to these categories in terms of correctness and speed. In the light of comparisons between categories, it was decided not to include zero as an added variable. Addition of an addend to a zero in the units position is much easier than considering the answer to a question which has numbers greater than zero in the units position.

(2) *Doubletons in both the questions and answers,* such as 33 or 55. These are visually eye catching and, as such, would have added another variable to the experiment. A doubleton can be typed in more quickly than two differing numbers as an answer. Doubletons were not included in the Key Search Reaction Time Test (see section 4.4).

(3) *Teen numbers* such as 14. When verbalising a teen number that the units number is 'said' first could cause confusion, especially for dyslexic participants when they are searching for the numbered keys to press in their answer.

(4) *Similar numbers in the tens or units* such as 53 + 3 or 26 + 27. These become easier if there is prior knowledge of the 2 times table. Moreover, this is an addition rather than a multiplication experiment.

(5) *Answers over 99*, thus eliminating any answer equal to or over 100. This decision limited the answer to two digits thereby avoiding the added complication of the 'hundred barrier'.

3.10.6 Question Order for Addition

Each question was assigned a number from 1 to 75. These numbers were placed individually in a hat and numbers were pulled out at random. Checks were made to see if there were any repetition of digits in the corresponding question in relation to the previous question, discarding such cases and returning the number back into the hat. Where an answer required was the same as a preceding question, that too was discarded.

The order of the questions is given in Appendix D (Table D.2).

Participants then reached the interval, which was of a variable length of time.

To continue with the second half of the Addition experiment, the participant needed to press the "Enter" key twice. This began Trial 2 in the one sitting, whereby all the questions were repeated but in reverse order. Thus the first question asked was '64 + 3 =' following through sequentially to the final question which was '21 + 7 ='.

At the end of each testing session the participants were thanked for their help but no comments were made on their performance.

3.10.7 Recording Components by the Computer

These are the same as for the Multiplication experiment (see section 3.8.6).

3.10.8 Response Categories for Analysis

These are the same as for the Multiplication experiment (see section 3.8.7).

Results for this experiment are given in Chapter 8 (Correctness) and Chapter 12 (Speed).

3.11 The Subtraction Experiment

3.11.1 Instructions for Program Working Disk (Main Instructions)

The instructions for the Subtraction experiment took the same form as for the Division experiment but the word 'subtraction' was substituted for 'division'.

3.11.2 Set Pre-test

As with addition this pre-test comprised 9 questions designed to evoke a response in ascending order of digits on the numeric keypad. The questions were as follows:

$$10-9,\ 10-8,\ 10-7,\ 10-6,\ 10-5,\ 10-4,\ 10-3,\ 10-2,\ 10-1$$

On completion of this pre-test the following instruction was displayed on the screen:

> Question ONE will start when you press "ENTER".

3.11.3 Random Pre-test

This test contained two questions from each of the five subtraction categories, outlined later. None of the questions would be duplicated in the Main Experiment.

Ten questions were asked in the following order:

$$27-2,\ 64-8,\ 81-3,\ 73-36,\ 41-7,\ 96-4,\ 79-6,\ 84-28,\ 58-7,\ 32-4$$

On completion of this pre-test the following instructions were displayed on the screen:

> You have completed the short section. [This was visible for only one second before the next instruction was given below.]
>
> Question ONE will start when you press "ENTER".

3.11.4 Main Experiment — Five Categories

As for the Addition experiment, *five categories* were arranged to follow the same pattern of questioning. The term 'subtrahend' is used for the number being subtracted.

Category 1 (for Subtraction)

This comprised low numbers (2, 3, 4) subtracted from two-digit numbers. The resulting answer did not cross the 'ten' barrier. It was labelled LN, which stands for 'Low subtrahend Not crossing the ten barrier'. Two examples of the LN category are: 29 – 2 and 67 – 4.

There were 15 questions in this category and each low subtrahend was used five times.

Category 2 (for Subtraction)

This comprised high numbers (6, 7, 8) subtracted from two-digit numbers. The resultant answer did not cross the 'ten' barrier. It was labelled HN, which stands for 'High subtrahend Not crossing the ten barrier'. Two examples of the HN category are: 87 – 6 and 58 – 6.

There were 15 questions in this category and each high subtrahend was used five times.

As a consequence of the design, all the answers include 1, 2 or 3 in their 'units'.

Category 3 (for Subtraction)

This comprised low numbers (2, 3, 4) subtracted from two-digit numbers and the resulting answer crossed the 'ten' barrier. It was labelled LC, which stands for 'Low subtrahend Crossing the ten barrier'. Two examples of the LC category are: 31 – 2 and 82 – 4.

There were 15 questions in this category and each low subtrahend was used five times.

Category 4 (for Subtraction)

This comprised high numbers (6, 7, 8) subtracted from two-digit numbers and the resulting answer crossed the 'ten' barrier. It was labelled HC, which stands for 'High subtrahend Crossing the ten barrier'. Two examples of the HC category are: 95 – 6 and 67 – 8.

There were 15 questions in this category and each high subtrahend was used five times.

Category 5 (for Subtraction)

This comprised two-digit numbers with high 'unit ' subtrahends of 6, 7 or 8 subtracted from two-digit numbers with the resultant answer crossing the 'ten' barrier with a 'borrowing' effect. It was labelled HARD, which stands for 'High subtrahend in the hard category'. Two examples of the HARD category are: 51 – 26 and 86 - 37.

There were 15 questions in this category and each high subtrahend was used five times.

Subtraction questions within each category are given in Appendix E (Table E.1).

3.11.5 Decisions about the Questions

Careful consideration was given to the digits used in the Subtraction experiment as in the Addition experiment. The experiment did not sample the whole population of subtraction sums and a decision was made to restrict the design to certain types of question otherwise the number of sums needed would have been too great. High (large) and low (small) subtrahends were selected as well as sums that crossed and did not cross the ten barrier because they seemed interesting; as a logical consequence, certain other sums were therefore excluded.

Decisions were made to exclude the following:

(1) *Any subtraction answer under ten* – this prevented one-digit answers. All answers were two digits for the purposes of comparison.
(2) *Subtraction of ten from any number.*
(3) *Negative numbers* – all answers were positive.
(4) *Taking away of doubles,* such as 80 – 40 or 16 – 8.
(5) *Taking away within a multiplication table,* for example 18 – 6 where 6 is a factor of 18.
(6) *Subtracting 0.* This has been avoided across all the experiments.
(7) *Subtraction of 'similars'*, such as 8 – 8.

Participants attended for this experiment and the Addition experiment on separate days so as to eliminate fatigue or confusion between the operations.

3.11.6 Question Order for Subtraction

A total of 75 questions were presented to the participant, which included a mixture of the five categories outlined above. The order of the questions was chosen as for the Addition experiment and the format was the same. See Appendix E (Table E.2) for the order of questions given in the Subtraction experiment.

Participants then reached the interval, which was of a variable length of time.

To continue with the second half of the Subtraction experiment, the participant needed to press the "Enter" key twice. This began Trial 2 in the one sitting, whereby all the questions were repeated but in reverse order. Thus the first question asked was '51 – 26 =' following through sequentially to the final question which was '52 – 7 ='.

At the end of each testing session the participants were thanked for their help but no comments were made on their performance.

3.11.7 Recording Components by the Computer

These are the same as for the Multiplication experiment (see section 3.8.6).

3.11.8 Response Categories for Analysis

These are the same as for the Multiplication experiment (see section 3.8.7).

Results for this experiment are given in Chapter 9 (Correctness) and Chapter 13 (Speed).

3.12 Statistical Tests

The following statistical tests have been conducted on the data.

3.12.1 Two-way Analysis of Variance

This test enables more than two samples to be studied simultaneously where differences between the means are compared. Variance is analysed within groups and between groups.

Once it has been established that differences exist among the means, post hoc range tests and pairwise multiple comparisons can determine which means differ. Range tests identify homogeneous subsets of means that are not different from each other. Pairwise multiple comparisons test the difference between each pair of means, and

yield a matrix where asterisks indicate significantly different group means at an alpha level of 0.05.

Tamhane's T2 test of differences is a multiple comparison test that does not assume equal variances.

3.12.2 General Linear Model (GLM) for Analysis of Variance

These are post hoc multiple comparison tests. Once it has been determined that differences exist among the means, post hoc range tests and pairwise multiple comparisons can determine which means differ. Comparisons are made on unadjusted values. These tests are used for fixed between-subjects factors only. In GLM Repeated Measures, these tests are not available if there are no between-subjects factors, and the post hoc multiple comparison tests are performed for the average across the levels of the within-subjects factors. For GLM Multivariate, the post hoc tests are performed for each dependent variable separately.

Tamhane's T2 test of differences is used when the variances are unequal (conservative pairwise comparisons test based on a t-Test).

The least significant difference (LSD) pairwise multiple comparison test is equivalent to multiple individual t-Tests between all pairs of groups. The disadvantage of this test is that no attempt is made to adjust the observed significance level for multiple comparisons.

3.12.3 Kolmogorov-Smirnov Z Test

This tests if two independent samples have been drawn from populations with the same distribution. Two-tailed tests are used throughout this thesis.

The Kolmogorov-Smirnov Z test is a more general test that detects differences in both the locations and the shapes of the distributions. The Kolmogorov-Smirnov test is based on the maximum absolute difference between the observed cumulative

distribution functions for both samples. When this difference is significantly large, the two distributions are considered different.

3.12.4 t-Test

This is a parametric test used to determine whether two means are significantly different from one another. Two-tailed tests have been used throughout the thesis.

The Independent-Samples t-Test procedure compares means for two groups of cases. Ideally, for this test, the subjects should be consistently assigned to two groups, so that any difference in response is due to the treatment (or lack of treatment) and not to other factors.

Statistics

In addition, for each variable: sample size, mean, standard deviation and standard error of the mean have been produced. For the difference in means: mean, standard error and confidence interval (the confidence level) can be specified. Tests: Levene's test for equality of variances and both pooled- and separate-variances t-Tests for equality of means have been produced.

3.12.5 Chi-square Goodness-of-Fit (Non-parametric)

This test compares observed frequencies with expected frequencies. It measures whether the distributions overlap or not. It is a test of how well a model fits the observed data. Small observed significance levels (say, less than 0.10) indicate that the model does not fit well.

3.12.6 Fisher's Exact Test

This is a non-parametric test that computes the exact probability of outcomes in a 2×2 table. It is used instead of Chi-squared where the observed frequency is 5 or less.

3.12.7 Mann-Whitney U Test

This test uses the ranks of the observations as a device for testing hypotheses about the identity of the two population distributions. Two-tailed tests are used throughout the thesis.

The Mann-Whitney U test is the most popular of the two-independent-samples tests. The observations from both groups are combined and ranked, with the average rank assigned in the case of ties. The number of ties should be small relative to the total number of observations. If the populations are identical in location, the ranks should be randomly mixed between the two samples. The number of times a score from group 1 precedes a score from group 2 and the number of times a score from group 2 precedes a score from group 1 are calculated. The Mann-Whitney U statistic is the smaller of these two numbers.

3.12.8 Spearman Rank-order Correlation Coefficient (rho)

This is a measure of the linear correlation between ranks. The closer the agreement between the rankings, the nearer rho is to 1.

The Bivariate Correlations procedure computes Spearman's rho and its significance levels. Correlations measure how variables or rank orders are related. Before calculating a correlation coefficient, the data is screened for outliers (which can cause misleading results) and evidence of a linear relationship.

Note

Significance levels are given as:

- $p < 0.05$ (significance at the 5% level) indicated with one asterisk (*);
- $p < 0.01$ (significance at the 1% level) and termed 'highly significant', indicated with a double asterisk (**);
- $p < 0.001$ (significance at the 0.1% level) and termed 'very highly significant', indicated with a treble asterisk (***);
- ns (not significant).

On a few occasions specific significance levels were reported where appropriate.

Reporting of results

The results for each of the mathematical operations are given in Chapters 6–13, which are further subdivided under two main headings: correctness and speed of response.

Chapter 14 gives details of 'instantaneous responding' across the four operations. Some 'additional behaviour' made by the participants was observed during testing and the findings are presented in Chapter 15. The performance of the dyslexics and non-dyslexics on the four operations is compared in Chapter 16.