Thesis

Submitted in part fulfillment of the requirements for the Continuing Professional Development Doctorate in Educational Psychology (DEdPsy)

Name of Course Member: Deborah Mary King

Title of Thesis: Types of questions asked by teachers in teaching numeracy to the whole class, and implications for children’s learning
Abstract
This was a study of the kinds of questions asked by teachers in the numeracy session. The aim of the study was to find out what sorts of questions are asked during the whole class teaching part of the daily mathematics lesson and whether conclusions can be drawn about the learning from these questions. A pilot study looked at questioning of children aged 8 and 9 (years 4 and 5) during the mathematics lesson, and a classification of question types was drawn up. Questions were categorised by raters into those which required factual recall, those requiring computation or numerical calculation, problem solving questions and class management questions. Raters agreed on classification of questions in the majority of cases.

In the main study teachers of children aged 8 and 9 were videoed in the whole class teaching section of the numeracy sessions, and their questions were classified into the four types above. To assess their learning in the introductory part of the Numeracy lesson children were given a short test based on the questions they had just been asked.

Teachers asked more computation questions than problem solving question and in three of the ten classes observed children performed better on the computation than the problem solving questions. Two types of question distribution emerged and these types of teaching style are discussed. The performance of the children in the post-test indicated that a number in each class had not understood the numerical concepts which had just been taught.
# Contents

<table>
<thead>
<tr>
<th>Contents</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>2</td>
</tr>
<tr>
<td>Contents</td>
<td>3</td>
</tr>
<tr>
<td>List of Figures</td>
<td>8</td>
</tr>
<tr>
<td>List of Tables</td>
<td>9</td>
</tr>
<tr>
<td>Chapter 1 Literature Review.</td>
<td></td>
</tr>
<tr>
<td>1.0 Introduction and rationale</td>
<td>10</td>
</tr>
<tr>
<td>1.1 Literature review</td>
<td>13</td>
</tr>
<tr>
<td>1.2 The development of numeracy in children</td>
<td>13</td>
</tr>
<tr>
<td>1.3 Children with difficulties in numeracy</td>
<td>23</td>
</tr>
<tr>
<td>1.4 Questioning in the numeracy session</td>
<td>27</td>
</tr>
<tr>
<td>1.4.1 Higher and lower level questioning</td>
<td>34</td>
</tr>
<tr>
<td>1.5 Conclusions</td>
<td>50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 2 Pilot Study: Classifying teachers’ questions in the numeracy lesson</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0 Abstract</td>
<td>53</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>53</td>
</tr>
<tr>
<td>2.1.1 Question Classification</td>
<td>53</td>
</tr>
<tr>
<td>2.1.2 Reliability and Validity</td>
<td>57</td>
</tr>
<tr>
<td>2.2 Method</td>
<td>58</td>
</tr>
<tr>
<td>2.2.1 Design</td>
<td>58</td>
</tr>
<tr>
<td>Chapter 3</td>
<td>Method</td>
</tr>
<tr>
<td>-----------</td>
<td>--------</td>
</tr>
<tr>
<td>3.0</td>
<td>Research Design</td>
</tr>
<tr>
<td>3.1</td>
<td>Participants</td>
</tr>
<tr>
<td>3.2</td>
<td>Procedure</td>
</tr>
<tr>
<td>3.3</td>
<td>Coding of questions</td>
</tr>
<tr>
<td>3.4</td>
<td>Videorecording as an observational method</td>
</tr>
<tr>
<td>3.5</td>
<td>Assessment of learning</td>
</tr>
<tr>
<td>3.6</td>
<td>Ethical considerations</td>
</tr>
<tr>
<td>3.7</td>
<td>Question types</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 4</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.0</td>
<td>Results</td>
</tr>
<tr>
<td>4.1</td>
<td>Part one: teachers' questions</td>
</tr>
<tr>
<td>4.1.1</td>
<td>Computation versus problem solving questions</td>
</tr>
<tr>
<td>4.1.2</td>
<td>Distribution of question types asked by teachers: teaching style</td>
</tr>
<tr>
<td>4.1.3</td>
<td>Relationship between gender of child and teachers' questions</td>
</tr>
<tr>
<td>4.2</td>
<td>Part 2: learning shown by analysis of post-test performance</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Proportion of children whose performance indicated difficulties</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Computation answers and problem solving answers</td>
</tr>
<tr>
<td>4.3</td>
<td>Relationship of questioning style to post-test scores</td>
</tr>
<tr>
<td>4.4</td>
<td>Scrutiny of errors</td>
</tr>
<tr>
<td>4.5</td>
<td>Research questions and conclusions</td>
</tr>
<tr>
<td>4.5.1</td>
<td>Part one: teachers' questioning</td>
</tr>
<tr>
<td>4.5.2</td>
<td>Part two: children's learning: post-test performance</td>
</tr>
</tbody>
</table>
List of Figures

Figure 2.1  Question classification by judges, showing the pattern of judges' classification of questions.
Figure 2.2  Percentages of question types.
Figure 4.1  Computation and problem solving questions by teacher as a percentage of total questions
Figure 4.2  Question Type Percentages for Each Teacher
Figure 4.3  Dendrogram of a cluster analysis showing two different groups by use of question type
Figure 4.4  Mean Percentages of question categories for type A and type B
Figure 4.5  Distribution of computation scores in T2 by gender.
Figure 4.5a  Distribution of problem solving scores T2 by gender.
Figure 4.6  Distribution of computation scores in T5 by gender.
Figure 4.6a  Distribution of problem solving scores T5 by gender.
Figure 4.7  Distribution of computation scores in T9 by gender.
Figure 4.7a  Distribution of problem solving scores T9 by gender.
Figure 4.8  Relationship between class management questions and correct post-test computation answers
### List of Tables

| Table 1.1 | Children's counting strategies. |
| Table 1.2 | Percentage of children performing an array task: children with developmental delay compared to a control group. |
| Table 1.3 | Classifications of mathematical questions. |
| Table 1.4 | Some classifications of general teacher questions (not specifically mathematics). |
| Table 1.5 | Higher and lower level questions. |
| Table 1.6 | Questions from the ORACLE study: percentages in 1976 and 1996. |
| Table 1.7 | Examples of the 6 question types from Perry et al. 1993, p.35. |
| Table 2.1 | Mean ratings for each type of question. |
| Table 2.2 | Mean question types to boys and girls |
| Table 4.1 | Means for computation and problem solving types of question. |
| Table 4.2 | Mean percentages and standard deviations for four question types by teacher groups A and B. |
| Table 4.3 | Mean percentages and standard deviations of question types to girls and boys. (i.e. percentage of total directed at girls and percentage of total directed at boys) |
| Table 4.4 | Percentage of children achieving below 25% in the post-test. |
| Table 4.5 | Percentage of children achieving zero in the post-test. |
| Table 4.6 | Mean percentage correct by teacher. |
| Table 4.7 | Classes with significant differences between computation and problem solving scores. |
| Table 4.8 | Teachers whose post-test scores did not show a significant difference. |
| Table 5.1 | Teachers' question types compared to the ORACLE findings |
| Table 5.2 | Differences in question classification. |
| Table 5.3 | Percentages of children scoring below 25% and scoring zero. |
Chapter 1

Literature Review

1.0 Introduction and rationale

In 1999 the Numeracy Strategy was introduced (Department for Education and Employment, 1999), a comprehensive mathematical curriculum for primary age children in England and Wales. The background to the introduction of the Numeracy Strategy was one of concern at the attainment of children in the United Kingdom compared with other countries. Comparisons such as the survey carried out by TIMSS, The International Mathematics and Science Surveys, (Keys et al., 1996 – 9) showed lower attainment in England and Wales than in Japan and other European countries. Although possible cultural factors should be taken into account, the differences were considerable; Butterworth (1999) pointed out that the mean score for the Iranian children was equal to the poorest 5% of the Singaporean (TIMSS, 1996).

The Numeracy Strategy is a national curriculum for primary age children which prescribes both content and, to some extent, the method of delivery or pedagogy. The strategy requires that a daily mathematics lesson is taught, broken down into three parts, the first devoted to teaching the whole class, the middle part to group activities related to the topic taught in the first part, and the last part to a plenary. One issue arising from this is how to differentiate the first part of the lesson so that it is accessible to all children. In early consultation on numeracy (DfEE, 1998) it was advised that differentiation be kept to a minimum in the light of concern about the British position in the national league tables for numeracy (Reynolds and Farrell, 1996; Keys et al., 1996). The Numeracy Strategy discussed differentiation within its framework in the introduction (DfEE, 1999, pp. 18-24).
It requested that staff ensure:

‘differentiation is manageable and centred around work common to all the pupils in a class with targeted, positive support to help those who have difficulties to keep up with their peers’ (DfEE, 1999 p.5). This desire to keep differentiation to a minimum may have implications for those children with difficulties in mathematics, estimated to be between 2% and 10%, (Butterworth, 2002; Snorre, 2002) and implications for children with language difficulties, given the medium of questioning is verbal.

Much of the teaching in this first part of the lesson is done through questioning. The Introduction to the Numeracy Strategy emphasised the importance of questioning, and offered guidance to good questioning (DfEE, 1999). While there has been considerable research on types of questioning in the classroom there has been less on questioning in mathematics, particularly in the relatively recent context of the Numeracy Strategy. It is of relevance to educational psychologists to consider questioning and its relationship to learning, and to consider the implications of whole class teaching for children with difficulties. While data is not available on referrals to educational psychologists arising from the Numeracy Strategy, there have been produced various ‘catch up’ programmes, such as ‘Springboard’ (DfEE, 2000) for children experiencing difficulties. It is possible, however, that certain parts of the numeracy session are inaccessible to a significant proportion of children.

This study set out to consider what types of questions were used in the numeracy lesson, how the children learned from them, and the implications of whole class teaching for children with difficulties.

In developing the research questions, studies from the following areas of research are reviewed
• cognitive processes in early arithmetic learning
• the relationship of arithmetical processing to language
• difficulties with the development of numerosity
• question types used by teachers during mathematics teaching. Studies of question types in numeracy will be discussed, particularly those in which it has been argued that there is a 'higher cognitive order' of thinking evoked by certain questions; the issue of, and underpinning rationale for such types of question and how these may relate to learning will be considered.
1.1 Literature review

To formulate specific research questions for this study, the literature review will firstly consider the broad issue of how children may develop numeracy skills and, in particular, the difficulties that some children have in developing these skills. Two different models of the cognitive processes underlying numeracy in children have been selected to provide a framework for the discussion of arithmetic development and learning difficulties specific to numeracy. The relationship of numeracy to language is considered as this is pertinent to the discussion of question types.

The second part of the review will focus on the questions asked of children in numeracy sessions in schools. Studies of question types in numeracy will be discussed, particularly those in which it has been argued that there is a 'higher cognitive order' of thinking evoked by certain questions; the methodology of these studies is considered.

1.2 The development of numeracy in children

In the study of the development of numeracy, researchers have defined the term as including two components, both the understanding of number as an abstract concept that can apply to any set, e.g. three cars, three planets, three ideas, three families, and the understanding of the relationships between them. This has been termed numerosity to encompass both the understanding of number and the relationships between numbers (see Butterworth, 1999; Nunes, 1996 for a more detailed discussion). Numerosity is more than numbers, 'numbers cannot be treated as straightforward representations of numerosity: they acquire meaning as part of a system that models relations between quantities' (Nunes, 1998, p.23)
It has been argued that children have a certain preparedness to develop numerosity. Butterworth (1999) talks of a numerical ‘start up kit’ including such things as the ability to count to about 4 or 5 without doing this sequentially. His evidence comes from cross cultural studies and work with infants and young children (for example, Wyn, 1992) indicating an innate awareness of numerosity. Understanding of concepts and mastery of skills may be triggered as children mature and learn from the environment. Whatever cognitive processes have been proposed in young children there does appear to be agreement that many children are often quite competent at calculation with small numbers by the age of 6. ‘It is well known that children can carry out numerical transformations involving addition and subtraction by 5 or 6 years of age’ [Huttenlocher, Jordan, & Levine, (1994), p. 284]

Interest from cognitive psychologists in how mental arithmetic problems are processed is relatively recent, (Ashcraft, 1992; Shalev et al., 1998) growing through the 1980s and 1990s during which time both an empirical base and theoretical perspectives were produced. Interest in the development and cognitive aspects of arithmetical processing has continued to grow (Butterworth, 1999; Baroody & Dowker, 2003). However, for the purposes of considering how children deal with the questions in the numeracy session, two models of arithmetical processing have been selected for consideration here, those proposed by Ashcraft (1992), and by Huttenlocher, et al., (1994). Such models discuss the architecture of processing in arithmetic and the development of numerosity in children and can be applied to how children may answer numeracy questions.

Ashcraft (1992) reviewed three models of arithmetical processing, his own, that of Siegler and Jenkins (1989), and Campbell (1987). As he argued that that the models were not radically different but presented progressive attempts to accommodate new empirical findings, only Ashcraft’s revised model will be discussed here. To provide
background it is useful to be aware that a prime focus has been the debate about whether answers are 'worked out' or retrieved from memory. Early work, (for example Groen and Parkman, 1972) assumed that children relied heavily on counting mechanisms, changing to retrieval mechanisms as they grew older. A child might need to count to add 7 and 3 while an adult was more likely to retrieve that number fact (Ashcraft & Fierman, 1982). However, both children and adults appeared to use strategies of varying complexity to work out answers, and further light was thrown on how children may process calculations from a study by Siegler (1987) who related the strategies used by 7 year olds to error rates and reaction times. These strategies appeared to form a hierarchy in terms of efficiency and sophistication.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Reaction Time (RT)</th>
<th>Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>retrieval</td>
<td>2.1s</td>
<td>4%</td>
</tr>
<tr>
<td>decomposition</td>
<td>4.1s</td>
<td>8%</td>
</tr>
<tr>
<td>minimum counting (min)</td>
<td>6.9s</td>
<td>17%</td>
</tr>
<tr>
<td>count all</td>
<td>16.3s</td>
<td>50%</td>
</tr>
</tbody>
</table>

(Siegler, 1987a, p.84)

Min. is used by the author to refer to adding on from the lowest number (see below)

**Table 1.1 Children's counting strategies**

Working from the simplest strategy, 'count all' is the system of counting from one in each case and it can be seen from this table this is inefficient and leads to a high error rate. 'min counting' is what in school is often called 'counting on' from a number, for example, given the sum, 8 add 7, the child starts with the larger number and counts on in units of one, 8, 9, 10 until s/he has counted on 7, often using
fingers, or other concrete aids. 'Decomposition' is breaking down numbers into smaller units, such as tens and units, to work with. For example 14 add 23 may become 10 add 20 and 4 add 3. However, beyond the age of 8 or 9, most processing for the basic facts of addition and multiplication is thought to be retrieval from memory (Ashcraft, 1992; Butterworth, 1999).

A model of arithmetical processing therefore should accommodate the use of retrieval of number facts, and the use of various working out strategies such as those above. However, Ashcraft pointed out that any model of arithmetical processing had to account for two consistent and robust empirical findings in mental arithmetic experiments. The first is the effect of number size on reaction times (RT). Children and adults are slower to answer questions in which the numbers are large than those with smaller numbers (Siegler, 1987; Geary & Wiley, 1991). Reaction times are longer, and errors are more frequent, with larger numbers and with zeros (Ashcraft & Christy, 1991). Ashcraft argued that this effect is related to experience; that is, how easy a problem was, depended on how often the child or adult had met it before. To prove that smaller numbers are experienced more often Ashcraft and Christy coded textbooks from grade one to six and found that smaller numbers appeared more often than larger ones and zero appeared least frequently of all. However, this effect of longer reaction times associated with larger numbers does not hold for what are known as tie problems, e.g. 2 + 2 or 7 + 7, where a relatively flat reaction time function has been found instead of the increase expected with the larger numbers. Frequently experienced computation problems, like high frequency words, appear to have a privileged position in the memory system for arithmetic.

Ashcraft went on to argue that the second finding, which had to be accommodated by any model, related to confusion, or types of errors and that certain errors indicated activation of related numbers. In experimental paradigms where the participant
decides whether an answer is correct or wrong, Ashcraft pointed out that there was increased difficulty in deciding whether the answer is right or wrong depending on how close the wrong choice is to the correct answer. In other words, the reaction time (RT) declined as the incorrect answer became more different from the correct answer but increased when the incorrect answer was close to, or related to the correct one. There was a longer RT in deciding whether answers such as $7 \times 4 = 21$ were correct than when the answer offered did not come from the same table – for example, $7 \times 4 = 18$, suggesting some sort of activation of a specific table. Furthermore, in adults and children, errors in multiplication were table mistakes, e.g. the answer to $9 \times 4$ was another multiple of either 4 or 9 in the majority of the errors, such as 27 or 32. Ashcraft suggested that a related but wrong answer resulted from some sort of activation and that the decision was then more difficult. What Ashcraft called the regularity and lawfulness of errors gave clues to how facts may be organised and stored (Ashcraft, 1992, p. 81).

Ashcraft argued that both the evidence from 'tie' problems ($7 + 7$), from reaction times, and from the errors of relationship discussed above indicated that related facts were in some way linked in storage. He has postulated a strong network of association similar to the model proposed by Collins and Loftus (1975). Collins and Loftus proposed that words which are related are grouped in connected clusters. For example car, lorry and bus may be linked and the strength of the links depends on the strength of the association. The faster reaction times to related words, and to very familiar words underlying this model have been found in more recent work; an example cited by Ashcraft was a study by Allen et al., (1992) producing convincing word frequency effects in a lexical decision task; participants were asked to respond quickly with a key press to a word or non word. Analogous to the proposal that the word fruit activated related words in the lexicon, and reaction times would be then faster to related words such as pear, Ashcraft argued that there would be activation
of numbers related to 3, such as 6 and 9. Such a spreading activation model would result in activation of related nodes. Children and adults would then use parallel routes using both retrieval and strategies to work out answers. Since associative networks become stronger with familiarity Ashcraft argued that adults would be more likely to reach the answer to familiar problems through retrieval, while children may use one of the strategies identified by Siegler (1987a) above.

This model is particularly relevant to some of the principles underlying the emphasis on practice of number facts in the numeracy hour. Spreading activation of the number network depends on familiarity, so that when older children and adults are asked a fairly familiar and 'easy' question such as 3 times 2, it is likely that the answer will be produced quickly by retrieval through a spreading activation network, without the use of strategies to find the answer. However with younger children, or adults faced with less familiar problems, there will be activation of strategies to work out the answer and it is likely that 'working out' the answer will not in that case be as fast as retrieval. Ashcraft's proposal of parallel routes is particularly relevant to young children where it may be supposed that strategies such as those identified above by Siegler (1987a) are used more often than retrieval in comparison with adults. Ashcraft argued that fact retrieval and strategy retrieval were triggered in parallel and the faster route would govern performance.

The relevance of this model to the emphasis on practice in the Numeracy Strategy (DfEE, 1999) lies in the theory that automaticity develops with repeated exposure to number facts, and this was echoed by Butterworth when he pointed out that much of the difference in competence with number was due to the memorisation of familiar calculations (Butterworth, 1999). Ashcraft's model allows us to consider that children may use various strategies until the familiar number facts have become a matter of retrieval.
However, two possible weaknesses in the relevance of this model are worth noting. Firstly the experimental paradigm used by Ashcraft (1992) was forced choice, in which the participant decides whether an answer was correct, for example, $9 \times 3 = 36$, where the participant makes a decision quickly. It could be argued that such a paradigm is less than ecologically sound. In most cases arithmetic problems are presented without an answer.

Secondly, the familiarity effect postulated by Ashcraft may depend in some way on the mediation of language; children may learn to say to themselves '60 add 40 is 100'. The relationship of language to the development of numeracy is complex and in children with difficulties in mathematics, it is difficult to disentangle the part played by difficulties with language processing and the part played by difficulty with numerosity (Levine et al., 1992; Jordan et al., 1992; Snorre, 2002). While it is not possible to explore this fully here the issue of language arises in relationship to the understanding, processing and output of the child presented with an arithmetical problem and is discussed more fully below.

The importance of the mediation of language arises particularly in considering how young children whose language system is not yet fully developed, may represent number. The second model considered here is that of Huttenlocher and colleagues (Huttenlocher et al., 1994), who investigated the part visual and verbal mediation might play, proposing an early internal visual model - a visual representation internally of the calculation. For example, a child may visualise the number three as an array of three dots, strokes, or sweets. Children may create this internal visual representation of number and, if asked to carry out a calculation such as one more or one less, they mentally add or take away one of the dots, strokes or sweets. In other
words they carry out a mental action to change their internal mental representation to help them reach an answer.

One way of investigating this has been to compare performance on visually mediated and verbally mediated number tasks (Levine et al., 1992). Children of 4 to 6 were asked to carry out calculation using a visual strategy or verbal strategies (Huttenlocher et al., 1994). To encourage the use of mental visual representation children were shown an array of discs which was then covered, and children asked to make a similar array. It was changed in front of the children by adding or removing one of the discs, covered again, and children were then required to copy what would be under the cover. Children were required to imagine the change. Children of 4 could do this with numbers less than 5 with great accuracy. However the success of children was less on a story problem where the same calculation was set in the context of a story, and less still in response to a number sentence, 'How much is 3 and 2'? On story problems they did not become proficient until they were older, and still older on number fact problems, where they did not become proficient until after 5. From these findings the authors speculated that children had a mental model for arithmetic before they learned the verbal skills and were able to use this preverbal, visual representation to work out arithmetical problems.

Huttenlocher and colleagues also hypothesised that that there might be social class differences evident in verbally presented arithmetic questions which might be less evident where a visual strategy was encouraged. The authors compared children from low income families with children from middle income families (United States) on the visual array method described above, on story problems and on number fact problems (Huttenlocher et al., 1994). The two groups of children did equally well on the array task (non verbal) but the lower socio-economic status children performed more poorly on both the number fact sentences and the story problems. The authors
concluded that children could solve non-verbal problems before verbally presented ones and that socio economic status was a factor affecting verbal strategies in arithmetic.

A further study by Huttenlocher et al., (1994) focused on children with learning difficulties, and the results once again raise the question of the role of language in children's number. They compared young children with developmental delay with a control group, on the disc array task. The table below shows their results:

<table>
<thead>
<tr>
<th>Numerosity</th>
<th>Control Group</th>
<th>Developmental Delay</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 disc array</td>
<td>94%</td>
<td>66%</td>
</tr>
<tr>
<td>4 disc array</td>
<td>52%</td>
<td>22%</td>
</tr>
<tr>
<td>5 disc array</td>
<td>34%</td>
<td>24%</td>
</tr>
</tbody>
</table>

(from Huttenlocher, Jordan & Levine, 1994, p.293)

Table 1.2 Percentage of children performing an array task: children with developmental delay compared to a control group

Quite apart from the difference between the two groups, it is interesting that even the control group had difficulties with the array of 5 discs, indicating perhaps that children's numerosity is less developed than might be assumed. Only 52% of the control group succeeded with the array matching of numbers up to 4. A comment made by the authors was that the differences shown in the table above on this non-verbal array task were actually greater than the differences between the two groups in language tasks, which included vocabulary and sentence comprehension. The placement of those children with learning difficulties had been based on language and social abilities. In this case, as language was a part of the screening, it might have been expected that the performance of the developmentally delayed children on the language task would be more affected than on the non-verbal array task, relative
to the control group. One possible explanation for the greater difference in the array task is that perhaps young children used language to mediate the processing of the array task and so those children with language difficulties were additionally disadvantaged. The question here is what the children actually did to transform the array of discs referred to above. Some may have talked themselves through it or reasoned aloud.

This raises the issue of the experimental demands of the array presentation which may have altered the way in which young children usually perform numerical calculations. It is worth noting that the use of an array may have encouraged the children to adopt this strategy, but it may equally be argued that the story problem encouraged both verbal and visual representation and the number fact problem encouraged verbal representation only. The researchers had supplied the mental model by showing the children the array, but it is not clear whether children would use this type of visual representation if it had not been supplied, nor in fact can it be concluded that this is what they did use. A more valid comparison might have been made between children taught using the array method and a number sentence method where reaction time could have been compared with both methods. However, as is the case with Ashcraft's model above we can see the influence of Huttenlocher's idea of internal visual representations of numbers in the visual aids of the Numeracy Strategy, (DfEE, 1999) such as the number line and the number grid.

From these studies, Huttenlocher and colleagues have suggested that children created a mental model and mapped the transformation onto this model – that is, they have seen the array of discs, they see another being added, or taken away, below a cover and they change their representation, or model. The design of the study attempted to remove the variable of language in early numerosity by using the
array task and demonstrated socio-economic differences in language-based computation which were absent in the array matching task.

It could be argued that both of the models discussed above worked from artificial paradigms and thus their relevance to learning mathematics in a school situation may be questioned. Much of Ashcraft's argument was supported by data from forced choice paradigms where the participant must choose the correct answer, and the data were the interference effects of the wrong choice slowing down the participants' decisions. Ashcraft's model postulated familiarity effects for numbers and groups of numbers similar to word frequency effects. It may be argued too, that Huttenlocher's paradigms predisposed the child to use an internal visual model by presenting and manipulating a visual array of objects. However, it is argued both models are relevant; Ashcraft considered familiarity and practice, a central theme of the Numeracy Strategy. Ashcraft's model predicted that if the questions were low in familiarity, children would be less likely to retrieve facts, and be more likely to use counting strategies. Huttenlocher's model pertained to the language of arithmetic and demonstrated the difficulty for children with number facts and number stories compared with visual representations, linking language processing and mental arithmetic. The models appear to be compatible and to have relevance to the consideration of children who experience difficulties with numeracy but the findings of Huttenlocher emphasise the part played by language processing in numeracy.

1.3 Children with difficulties in numeracy

To some extent arguments presented by Snorre (2002) have bridged the gap between the two models developed by Ashcraft (1992) and Huttenlocher, Jordan and Levine (1994) and described in the previous section. Snorre (2002) has argued that
some children with linguistic difficulties are 'mathematically delayed' and catch up; others are 'mathematically disabled' and this group continues to have mathematical difficulties. Those with delay show developmental differences in what he termed arithmetical word problems, as they have to integrate linguistic processing with arithmetical processing (Snorre, 2002). He compared children's performance on problems such as: 'Peter has seven marbles and Sam gives him eight. How many does he have altogether?' with the number sentence $7+8 = ?$ and found that 'mathematically delayed' children were more affected by the word form of the question the younger they were. In other words, the language had an effect on their understanding. However, these children became less affected by the form of the question as they grew older and were then able to do the mathematics in the question.

On the other hand, children he termed 'mathematically disabled' were less affected by the form of the question and had relatively greater difficulty with the mathematical content. They made only slight progress, reaching a peak at grade 2 (7 years old) while the other group made marked progress over the primary years. Thus the 'disabled' group continued to find the mathematics difficult, while the 'delayed' group 'caught up'. Such an argument appears to attribute the difficulties at least partly to language in number questions. Snorre argued that most children learned to use mental strategies increasingly over the years while those with mathematical difficulties ('mathematically disabled') continued to need concrete counting strategies.

Butterworth (2002) suggested that there were between 3% and 9% of children in the United Kingdom with dyscalculia. He defined dyscalculia as 'blindness to numerosities' (Butterworth, 1999, p.265) but later he made it clear that by dyscalculia he referred to a specific difficulty with the relationships between numbers. He has suggested this may be linked with specific damage to the interparietal area of the
brain (Butterworth 2002). Butterworth has recently developed a screening system, proposing that three tasks differentiate children with dyscalculia from the others. The first is a Stroop type number task in which the child is asked, ‘which is the greater number – 3 or 6?’.

Children with mathematical difficulties find it much harder to give the correct answer if the lesser number is physically larger. The second task is what is known as subitizing, or estimating the number of dots in an array. The third task is timed arithmetic where a simple measure of reaction time differentiates between children with mathematics difficulties and those without (Butterworth, 2002). A pertinent question is whether his screening would differentiate between the two groups of children discussed by Snorre (2002) who were either mathematically disabled or mathematically delayed. The implications of Butterworth’s estimate of the percentage of children with dyscalculia are that in any class between one and three children may be found who would have such difficulties.

A lower estimate comes from a longitudinal study of children with arithmetic difficulties carried out by Shalev et al., (1998) who argued that between 2% and 6% of children in mainstream have difficulties with arithmetic which would not be predicted from their cognitive ability. In common with Butterworth, (1999) Shalev and colleagues defined dyscalculia as the presence of a specific difficulty with numbers and carried out cognitive assessments to exclude children with more general cognitive difficulties. They carried out a longitudinal study of children of 9 and 10 years of age, first screening a group of over 3000 and following them up at 13 and 14 years old to assess the persistence of difficulties with arithmetic. Dyscalculia was defined as difficulties with arithmetic which placed children in the lowest 5% of their sample, but excluding those whose general cognitive scores were below average (below the 80th centile). They defined two groups of children at age 13, those with
persistent dyscalculia who remained in the lowest 5% of those tested on an arithmetic battery despite school interventions, and those with less marked difficulties, whose scores fell in the lowest quartile at age 13. The most marked difference between the two groups was in division (none of the lowest 5% had mastered it, compared with 56% in the lowest quartile). Complex subtraction also differentiated markedly between the lowest quartile and the lowest 5%; only one fifth of those defined as dyscalculic, the lowest 5%, had mastered complex subtraction at the age of 14. Complex subtraction requires the decomposing of tens into units or hundreds into tens. The fact that this study was both longitudinal and large scale lends weight to the authors' estimate of 'persistent developmental dyscalculia' in around 2-3% of their original sample. These children despite intervention continued to score in the lowest 5% and were more likely to show attentional and behavioural difficulties.

There were unfortunately gaps in the data; over 25% of the original children screened were not available for follow up; there was no data about the type, frequency or length of the interventions, defined by an individual arithmetic tutorial; no data from teachers were available about either interventions or behaviour, nor were data gathered on other difficulties such as reading or language. Information on behaviour and attention was gained from parents only.

Shalev et al., (1998) argued that these figures were comparable with those of Rutter and colleagues in the Isle of Wight study for the persistence of dyslexia (Rutter, Tizard, Yule & Whitmore, 1976). If we add to this estimate of children with specific number difficulties those children with more general learning difficulties including number difficulties, it may be argued that as many as 10% of the class might have significant difficulty with the content of the numeracy session and specifically with the whole class session.
To sum up, such arguments indicate that a significant proportion of primary age children may have difficulties with the learning in the numeracy lesson. The findings of the effect of language on computation (Huttenlocher et al., 1994; Snorre, 2002) would predict that more children would have difficulties with problem solving questions than with computation. The estimations of Butterworth (2002), Snorre (2002) and Shalev et al., (1998) would predict that between 2% and 9% of children would have difficulties with the questioning in the whole class teaching part of the numeracy session.

1.4 Questioning in the numeracy session

Research discussed in the previous section points to the difficulties which may arise for some children from the questions asked in the numeracy session. It has been argued that there is a significant proportion of children with numeracy difficulties and that there may be an additional group with language difficulties; both groups are likely to find the questioning in the first part of the numeracy lesson difficult. The Numeracy Strategy has emphasised the importance of questioning as a vehicle for both assessment of children's knowledge and learning (DfEE, 1999). The Office for Standards in Education has indicated a strong association between the quality of teaching and the standards achieved in various subjects; the most significant aspect of pedagogy was the teachers' ability to use questions: ‘effectively to assess the pupils' knowledge and challenge their thinking’ (Ofsted 1994, 1995).

Narrowing the focus to mathematical questioning, various researchers have used classifications of questions. For this study the researcher has attempted to show schematically the use of question typing by various authors. A brief overview of
some of these question classifications is presented in the tables below; Table 1.3
shows some classifications of question types in mathematics, and Table 1.4 shows
some classifications of more general teacher questions. Studies by Perry,
Vanderstoep, and Yu, (1993) and by Newstead, (1998) are included because they
have classified the questions asked by teachers in numeracy, and carried out
empirical research using their classification system. As the Numeracy Strategy is the
subject of this study, classifications from the DfEE (DfEE, 1999) and from Ofsted
(Ofsted, 2000) are also included, although it does not appear that these are based on
empirical research. It can be seen, however, that there are common threads in the
identification of certain question types.

The table below, Table 1.3, has been drawn up to show that question types in
mathematics teaching appear to fall into three categories; these are shown by colour
outline. Open questions, often described as questions requiring higher level
reasoning, are boxed in yellow. Closed questions which have one correct answer
are boxed in red. Questions to do with class management and organization are
outlined in blue. An x indicates use of the description by the author or authors in the
right hand column. Table 1.4 applies to studies of more general questions in
teaching, but uses the same colour classification.
<table>
<thead>
<tr>
<th>Type</th>
<th>Level</th>
<th>Author</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High Level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Open</td>
<td>Askew &amp; William, 1995</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low Level</td>
<td>Boaler, 1997</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low Level</td>
<td>DfEE, 1999</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low Level</td>
<td>Kerslake, 1982</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low Level</td>
<td>Newstead, 1998</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low Level</td>
<td>Ofsted 2000 Review of Numeracy Strategy</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low Level</td>
<td>Perry et al, 1993</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low Level</td>
<td>Shoulders &amp; Birley, 1993</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low Level</td>
<td>Leech &amp; Babington, 1998</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Closed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1.4 Some classifications of general teacher questions (not specifically mathematics)

<table>
<thead>
<tr>
<th>Author</th>
<th>Open</th>
<th>Closed</th>
<th>Closed Low Level</th>
<th>Closed Low Level</th>
<th>Open/Closed High Level</th>
<th>Closed Low Level</th>
<th>Closed Low Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Galton et al, 1999</td>
<td>✓</td>
<td>X</td>
<td>fact recall</td>
<td>recall of fact</td>
<td>open solution</td>
<td>recall</td>
<td>shape understanding</td>
</tr>
<tr>
<td>Kyriacou, 1995</td>
<td>✓</td>
<td>X</td>
<td>recall of fact</td>
<td>recall</td>
<td>open solution</td>
<td>recall</td>
<td>shape understanding</td>
</tr>
<tr>
<td>Mills, 1996</td>
<td></td>
<td></td>
<td>recall</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morgan &amp; Saxton, 1994</td>
<td></td>
<td></td>
<td>elicit information</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In both the mathematics and the more general classifications, there are commonalities, such as whether the question is open or closed. In the mathematics classifications there are further commonalities such as whether the question demands a factual answer (column four), whether it demands a calculation (column five), or sets a problem to be solved (column six). The top row refers to the fact that some authors have also discussed questions in terms of whether they evoke high or lower levels of thinking and this will be discussed later in this chapter.

In Table 1.3 the broadest category in columns two and three is the open and closed dichotomy. Where there are crosses, these authors use this categorisation. In general terms, an open question is one which has more than one possible answer (Brown & Wragg 1993; Kyriacou, 1995; Mills, 1996; Galton et al., 1998; Ofsted., 2000). The next columns, four to six, refer to more specific types of mathematical questions. The first category, in column four, is those questions which tap factual observation or factual recall, and are closed questions with one answer. Column five, computation, is also a closed type of question, requiring calculation of an arithmetical operation. This is in contrast to problem solving questions in column six which may be closed or open questions with more than one answer. Finally, in column seven there is a group of questions which relate to management of the class, organising groups and tasks, and maintaining the focus of the children; this category is called ‘class management’ for the purpose of this study.

The following are the categories with examples. Wherever possible examples have been taken from the authors but in some cases the authors have defined a question type without examples (Boaler, 1997; Newstead, 1998) and for the sake of clarity, some examples are given which have been taken from class observations made prior to this study.
open:  'What do you think is happening when the solution turns blue?'  
(Galton et al., 1999, p.29)

closed:  'Well, is it 4 or 5?'  (Boaler, 1997, p.22)

factual recall:  'Is this number in the ones column or in the tens column?'  
(Perry et al, 1993, p.34)

‘What is the name of this shape?’  (DfEE., 1999)

computation:  ‘What is 2 x 3?’  (Kerslake, 1982)

problem solving, including:

- selection of operation, or computation in context:  ‘If there are 20 teams in a football league and each team has to play every other team, how many games will there be in all?’  (DfEE, 1999, p.9) and  ‘There are 30 pieces of red folding paper, 20 pieces of green. How many altogether?’  (Perry et al., 1993 p.34)

- explanation of concepts:  ‘Why did you use subtraction for this problem?’  
(Perry et al., p.34,1993)

- hypothesis formation:  ‘From this information what relationship has foot size to height?’

- enquiry  ‘What do you think of that?’  ‘Can anyone think of an easier or shorter way?’

class control:  ‘Are we ready to go on again?’  (Morgan & Saxton, 1994)
While the table attempts to impose a systematic classification there are three significant difficulties with the ways in which various researchers have categorised question types. The first is the confusion between factual recall and computation, the second is the range of different terminology and descriptors which can be misleading, and the third lies in the range covered by problem solving questions.

The confusion between factual recall and computation has arisen from systems such as Kerslake's (1982), who referred to a class of question as factual recall, with the example of '2 times 3'. In other systems this might be classified as computation (Askew & Wiliam, 1995). Perry et al., (1993) have combined rote recall and computation in a 'factual recall' category, while they combined rule recall and computation in a 'computation' category. The difficulty in separating out factual recall from computation is that we do not know what cognitive processing was done by the child to answer the question. If a child calculates the answer to '2 times 3' then the category would be computation; if the child retrieves the number fact from memory then the question is factual recall. It is relevant here to return to Ashcraft's model (Ashcraft, 1992). He proposed that both retrieval of a number fact and working out the answer would work in parallel, with the fastest route producing the answer. So for a child who is familiar with this number fact the question may be one of recall, and for a child who is not familiar with it then it is more likely to be a question of computation or working out.

This signposts the second difficulty which is that there are two very important unknown variables in any question classification system. The first is the teacher's intention in asking the question; although it has been suggested the teacher ask, 'What do I want this question to do?' (Morgan and Saxton, 1994, p.41), an observer cannot know the intention of the teacher. The second is what the student does to
answer the question. It must be acknowledged that any classification system may not reflect what the child does with a question nor the intention of the teacher; this creates difficulties for the construct validity of any classification system, and is further discussed in Chapter 3, Section 2.4 when discussing how children reach an answer (Campbell and Fuselgang, 2001), and in more detail in Chapter 6, Section 2.

The third difficulty is clear from the range of descriptors used in the problem solving category. This category covers a wide range of questions, from selection of the correct rule or operation to hypothesis formation and enquiry (Boaler, 1997; Kerslake, 1982; Newstead, 1998; Ofsted, 2000; Perry et al., 1993). It is further complicated by the term, computation in context (Perry et al., 1993). An example of computation in context might be, ‘How many sweets will we need to give each person in the group three?’ As this requires the selection of the correct operation, addition, subtraction, multiplication or division, this would be defined by others as problem solving. The range of question types included in the problem solving appears to be defined by mathematical questions which require something other than recall or working out. As all of such questions appear to require more of the child than computation alone, the generic term of problem solving appears appropriate.

1.4.1 Higher and lower level questioning

Some researchers have made the assumption that certain questions are in some way more likely to evoke what has been called higher order mathematical thinking in children (Boaler, 1997; Galton et al., 1999; Kyraciou, 1995; Perry et al., 1993; Ofsted, 2000). This distinction is echoed in an early review of the Numeracy Strategy (Ofsted, 2000) which drew attention to the different types of question. The report pointed out that this was done well in only six out of ten lessons, where there had been:
'a good range of open and closed questions' but 'where the teaching was unsatisfactory too much time was spent on the recall of number facts rather than expecting pupils to figure out new facts from known ones and explain how they had arrived at their answers.' (Ofsted, 2000, p.6)

Looking at Table 1.5, it can be seen that the terms used to refer to higher and lower level categories correspond with terms used to refer to open and closed questioning. There appears to be no evidence of theoretical rationale or model underpinning such categorisation however, and this will be further explored in this section.

<table>
<thead>
<tr>
<th>Author</th>
<th>Level</th>
<th>Low Level</th>
<th>High Level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>computation</td>
<td>problem solving</td>
</tr>
<tr>
<td>Kyriacou, 1995</td>
<td></td>
<td>rule recall</td>
<td>strategies</td>
</tr>
<tr>
<td>Perry et al, 1993</td>
<td>computation</td>
<td>rote recall</td>
<td>general problem</td>
</tr>
<tr>
<td>Newstead, 1998</td>
<td></td>
<td>product question</td>
<td>solving</td>
</tr>
<tr>
<td>Boaler, 1997</td>
<td></td>
<td>procedural</td>
<td>process question</td>
</tr>
<tr>
<td>Galton et al, 1999</td>
<td>closed fact</td>
<td></td>
<td>conceptual knowledge</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>open, 2 or more</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>answers</td>
</tr>
</tbody>
</table>

Table 1.5 Higher and lower level questions

Boaler (1997) argued that any question which could be answered by the application of a rule, method or formula was procedural, requiring a lower level of thought while questions requiring mathematical thought were conceptual, requiring a higher level of thinking. Although Boaler did not define mathematical thought she indicated that conceptual questions 'required a depth of thought which would be useful in a number
of applied and real world settings’ (Boaler, 1997, p.81). She offered as an example, 
*A shape is made up of four rectangles, it has an area of 220 cm². Write in terms of x, the area of one of the rectangles’* (Boaler, 1997, p.77). If a question could be answered by memory alone it was procedural, if a question needed thought as well, or instead, it was conceptual. Askew and Wiliam (1995) described lower level questions as those concerned with fact or procedure, while higher order questions demand the application of principles, synthesis, and explanation. Their example of a lower order question was:

‘*What is the name of this shape?’* while their example of a higher order question was, ‘*How would you describe this shape on the phone to a friend so he could draw a copy?’* (Askew and Wiliam, 1995, p.16). Other research has referred to these ‘higher and lower’ level questions as product and process (Newstead, 1998). Questions were two types, product questions, ‘*which elicit short, simple predictable answers*’ and process questions ‘*which are broader, eliciting more expanded thinking, e.g. why or how*’ (Newstead, 1998, p.59).

While it is easy to see why the assumptions of high and low level thinking may have been made, it is difficult to define what various researchers have meant by this or to gain a clear idea of what cognitive processes may be involved. Three studies are discussed, and the concept of higher order questioning in these studies is explored. The first and most general study is a large scale study of teacher - pupil interaction (Galton et al., 1998). The second is a cross cultural study of teachers’ mathematics questions, which aimed to explain cross cultural differences in achievement (Perry et al., 1993). The third attempted to categorise the teachers, at least partly on their questions, into traditional and reform (Newstead, 1998), and is included because it recognised a social anxiety element in answering mathematics questions.
The first study is included here because it is the second of two large scale observational studies of interactions in class, separated by twenty years, one in 1976 and one in 1996, (Galton, Hargreaves, Comber, Wall, & Pell, 1999). These studies were known as the ORACLE studies (Observation, Research, and Classroom Learning Evaluation). From 1976 to 1978, fifty-eight teachers were observed for three days each term, and the pattern of both teachers' questions and statements were analysed. Twenty years later, Galton and his colleagues attempted to replicate this, sampling fifty-eight teachers of years five and six in primary school, that is, teachers of ten and eleven year old children. They looked at, among other things, types of questions and whether questions were directed at boys or girls (Galton et al., 1999). Questions were categorised into factual questions, open and closed questions, task supervision and routine questions (see Table 1.6).

<table>
<thead>
<tr>
<th></th>
<th>1976</th>
<th>1996</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Percentage (%) of all Questions</td>
<td>Percentage (%) of all Questions</td>
</tr>
<tr>
<td>of fact</td>
<td>24.7</td>
<td>24.7</td>
</tr>
<tr>
<td>closed solution</td>
<td>18.3</td>
<td>34.6</td>
</tr>
<tr>
<td>open solution</td>
<td>5.0</td>
<td>9.9</td>
</tr>
<tr>
<td>task supervision</td>
<td>32.5</td>
<td>18.5</td>
</tr>
<tr>
<td>routine</td>
<td>15.0</td>
<td>12.3</td>
</tr>
</tbody>
</table>

(taken from Galton, Hargreaves, Comber, Wall, Pell, 1999, p.30)

Table 1.6 Questions from the ORACLE study: percentages in 1976 and 1996

In the original ORACLE study it was noted that teachers: 'who engaged in above average whole class activity were more likely to be engaged in task related interactions, including the highest proportion of what were termed higher order or challenging questions' (Galton et al., 1999, p.27), and the authors noted that such
results were used to justify the use of whole class teaching for enquiry; hence the emphasis on whole class teaching in the Literacy and Numeracy Strategy (DfEE, 1997). However the authors argued that the shift to whole class teaching over the twenty years had resulted in what they called an increase in,

: ‘talking at pupils through statements and not in talking with pupils through asking questions’ (Galton et al., 1999, p.27; original emphasis retained).

The type of questions asked were of a, ‘low cognitive level, requiring one or two-word responses, and many were rhetorical’ (ibid. p.28)

The authors argued that over the twenty years:

• changes were quite small apart from a drift towards more closed questions in whole class teaching and more talking at the class. In 1976 non interaction, housekeeping activities took up over 20% of the day, while in 1996 these activities took up far less class time.

• questions were about 1/3 of statements. The percentage of what the authors defined as closed questions had doubled in 1996; these were questions with one possible answer, although the authors made the point that this was difficult to define. (It appears that the authors have not considered factual questions as a subdivision of closed questions; if these are combined it appears that 60% of questions are closed, excluding those concerned with task supervision and routine).

• teacher pupil interaction was greatest in a whole class setting (such as the literacy or numeracy lessons).

• interaction had increased and this increase was because the teacher talked more at the class – ‘an increase in talking at pupils through statements and not in talking with pupils through questions’ (Galton et al., 1999, p.27)

They noted that little had changed in the pattern of teacher interaction over the twenty years and that the questions were, in the majority, undemanding:
'open or speculative or challenging questions - - are still comparatively rare'
(Galton et al., 1999, p.33)

The classification of questions in this study raises several issues. The first difficulty lies with the classification of closed and open questions; these appear to be superordinate categories, into which other types of question, such as questions of fact, would fit. A second difficulty in analysing question types is related to the difficulty of identifying what the teacher’s intentions were in asking a question; this point was raised by the authors so they adopted a classification based on the teacher’s response to the answer offered by the pupil. If the teacher actively sought or accepted more than one answer then the question was defined as open. The authors admitted that such recording did not take account of the questions where no pupils offered alternative answers.

Thirdly, such a definition – open and closed – presented difficulties as an instrument for use in numeracy research. Teachers may develop a closed question into an open question. An example is a Japanese teacher placing 5 rows of 10 tiles on the desk, taking away three rows, asking how many are there, and following the answer asking the class how they knew it was a subtraction problem (Perry et al., 1993). Teachers may use a wrong answer to a closed question to explore a common misconception (William, 1999) using either a series of closed or open questions.

Fourthly, and perhaps most importantly, there appeared to be an assumption of ‘higher order’ questions by Galton and his colleagues. A question to which a teacher expected more than one answer was inherently a ‘better’ question, of a higher cognitive order. The authors did not clarify why they supposed such questions to be of higher cognitive order, nor indeed to define what a higher cognitive order might be. It might be argued that carrying out a series of calculations demanded by multiplication of two three figure numbers is of a higher cognitive order than a
question asking a child to explain how s/he might check the answer to 3 times 4. It was assumed by Galton and his colleagues that a question demanding more than one answer, and intended by the teacher to demand more than one answer, is in some way better than a question which is intended to have only one answer. Overall, the approach to the classification of questions appeared arbitrary and the assignation of greater cognitive value to certain question types appeared to have little theoretical rationale. The authors have themselves acknowledged the difficulties in categorising questions pointing out that the results in individual categories need to be interpreted with some caution. However, they argued that the increase in open and closed questions was achieved largely by the decrease in task supervision questioning. While the categories of teachers' statements are not pertinent here, the data concerning statements and questions may well not be independent; an increase in one type of statement may be at the expense of one type of question. The data were further complicated by the inclusion of some types of question in the statement classification – rhetorical questions.

Scarath and Hammersley (1986) criticised the methodology of the first ORACLE study, in particular the coding of questions, arguing that reliability may have been poor on some question categories, although details of reliability were not provided by the ORACLE study, apart from an overall figure of 81% on inter observer agreement. As Scarath and Hammersley pointed out, on some areas such as 'Silence' one could reasonably expect 100% agreement. This indicated some differences in observer coding of questions and statements. They have also drawn attention to the difficulties in coding a question by how the child answered. For example, the category of recall questions was open to debate; it may be recall, or it may be something that the pupils worked out from information they had (see the discussion on recall and working out above). Furthermore, they criticised the sampling and the validity of the research, pointing out that the coding rules were not sufficiently
unambiguous to be sure that different observers would arrive at the same code for a statement or question.

Given these reservations about the data, it may be reasonable to query some of the conclusions of the ORACLE study. Galton and colleagues argued that the changes over twenty years were not large: *today's primary teachers - - - - have continued to engage in the same pattern of exchanges with their pupils as their predecessors did two decades ago* (Galton et al., 1999, p.34). Teachers, they argued, were in danger of becoming technicians, and that, *teaching the class, as this study shows, is rarely stimulating* (ibid p.34)

Although the ORACLE study concluded that there had been little change in the nature of interactions over this twenty year period, the authors argued that pedagogy lies in selecting methods or teaching strategies which are fit for the purpose and that the job of trainers was to advise on such strategies. The authors speculated on a change in the attitude of teachers, who may not see the art of pedagogy as a choice of method to suit each subject. It is unclear why, given their own conclusion that there has been little change, the authors should argue for this 'deprofessionalisation' in teachers' attitudes. The argument that pedagogy cannot be decontextualised, may be applied to the difficulties with the ORACLE observations and data analysis. The data and the conclusions were decontextualised. Finally, if, as the authors argued, the judgement of teaching methods lies in the its effectiveness, or how well it enables children to learn the subject of the teaching, the ORACLE study appears to have lacked a measure of effectiveness provided by an assessment of children's learning.

In contrast to this, the second study of mathematical questioning which is selected for discussion took a starting point of differences in achievement (Perry, Vanderstoep & Yu, 1993) following up an interesting cross-cultural comparison reported by Stigler
Lee and Stevenson, (1990); Stigler and colleagues found that children in the United States in 1st and 5th grade (6 years and 10 years of age) performed more poorly than Japanese and Taiwanese children, on a range of mathematical questions, both computational, that is straightforward calculation, and problem solving. These differences were larger for the older children than for the younger, indicating that school did nothing to decrease the gap. Because the disparity increased with the age of the children, it was hypothesised that variation in teaching practices might account for the increasing difference. The difference at the start of school may be more difficult to account for, except by an assumption of cultural variation.

Following this, a cross cultural analysis of question types was carried out (Perry et al., 1993). They selected ten schools in Japan and Taiwan and 20 in the US, using 16 observers in Japan, 10 in Taiwan and 4 in the US. (There was no information on how the schools were selected). In each school two first grade (children of 6-7) class teachers were observed. These observers did not use video recorders nor tape recorders but were told to include as many verbal remarks as possible and to describe behaviour, materials, and the specific nature of the content. Summaries of these observations were produced – though it is not clear whether the observers produced their own summaries, or whether they were produced by researchers, and the summaries were coded by two researchers, using a coding scheme developed to classify the questions by the authors. The classification into question types was as follows:

<table>
<thead>
<tr>
<th>Question Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>computation/rote recall</td>
<td>7+5 =</td>
</tr>
<tr>
<td>rule recall</td>
<td>What is the rule for 2 digit addition?</td>
</tr>
<tr>
<td>computing in context</td>
<td>There are 30 pieces of red paper, 20 green</td>
</tr>
</tbody>
</table>
Table 1.7  Examples of the 6 question types from Perry et al., 1993, p.35

<table>
<thead>
<tr>
<th>Problem Type</th>
<th>Question Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>generate a problem</td>
<td>Make a problem around 8-3-2=</td>
</tr>
<tr>
<td>problem solving strategies</td>
<td>How did you arrive at the answer?</td>
</tr>
</tbody>
</table>

The results of the observation indicated that, across the three countries, there was no significant difference in 'computation/rote recall,' no significant difference in 'rule recall,' and no significant difference in children 'making up a problem'. Significant differences were found in 'computation in context'; there were more of these questions in Japan and Taiwan, and in such questions the authors said the context was less meaningful in the United States, although they did not explain what criteria were used to judge meaningfulness. Significantly more 'problem solving strategy' questions were asked in Japan and Taiwan.

There appear to be some difficulties with this classification. The authors have said they chose deliberately to classify questions not on responses, but on teachers’ intentions, although it is not clear how these were established. The issue of validity arises with the question types devised by the first two authors; no rationale was given for the types described. For example, it seems curious to have as one category rote recall together with computation, while a researcher such as Ashcraft (1992) discussed above has typically separated computation and rote recall. Other research does not indicate that first graders (children of 6) would be carrying out computation by rote recall. Indeed Huttenlocher found only 34% of five year old children could perform calculations with numbers below 5 (Huttenlocher, Jordan & Levine, 1994). The validity of such question types is not made clear, nor how they were derived.
Secondly, the authors concluded that the questions asked of Asian children were more challenging than those asked of children in the United States. The authors have assumed that the conceptual questions demanded a 'higher level' of thought than those asking the child to generate a problem, but offered no rationale for this. ‘We believe that asking higher level questions may lead to the formation of the relatively rich and broad conceptual knowledge about mathematics demonstrated by Asian children’ (Perry et al., 1993 p.39). Despite this they themselves warned against a causal hypothesis, i.e. that asking a certain kind of question produced a certain effect.

Thirdly, there were issues about the methodology; there were 30 different observers and no evidence was gathered such as tape recordings or video recordings to corroborate their observation; no interrater reliability was reported, although all lessons were coded by two coders and any disagreements resolved by discussion among the whole group of coders. Observers were told: ‘to pay special attention to the verbal remarks made by the teachers and students and to include as many of these as possible in their notes’ (Perry et al.,1993, p.33)

There was further opportunity for distortion of the observation in that summaries were made; the authors did not explain why these were necessary, nor who did the summarising, and how decisions were made to omit observational data. Finally, it was unclear whether these questions were proportions of all the lessons taught or just the mathematics lessons.

On several grounds, then, particularly those of reliability and validity, the study made it difficult to generalise, as the authors have done, to other subjects and age groups, claiming certain types of questions encourage what they call higher order thinking.
What might account for the difference in the attainment of Japanese, Taiwanese and American children (Stigler et al., 1990; Perry et al., 1993)? It may be due to the use of different types of questioning but the following variables noted by Stigler and colleagues may also have contributed to part of the variance. There was more public evaluation of the children in Japan and Taiwan; in the US the evaluation was typically private. The Asian children spent longer in the classroom with a longer day and fewer days holiday. The U.S textbooks covered fewer topics and U.S children had covered only 66% of these topics by fifth grade compared with 86% covered in Japan (Stigler et al., 1990).

Finally, it should be remembered that such achievement differences might have been partly due to cultural child rearing variations. It is interesting that a difference between the children exists in first grade, and the possibility should be considered that the difference exists because of home influences and may increase for the same reason. There may be less emphasis on arithmetic and numeracy in some cultures, and a variation in the part that home expects to play in the learning of the child.

The final study of questioning to be considered here has classified the teachers into two groups at least partly on their mathematics questioning style, and once again a certain type of question has been considered to evoke a higher level of thinking (Newstead, 1998). Part of the observation of teachers in this study was the Verbal Interaction analysis, which, Newstead argued, showed how far the teacher was, ‘giving problems or concepts rather than giving facts’ (Newstead, 1997, p.59) and how important the teacher considered higher level thinking, that is, explanation and justification, in addition to factual information. The proportion of teacher questions which were process questions was one of the criteria for discriminating whether
teachers were 'traditional' or 'alternative'; she examined the proportion of questions which were factual and computation but she also used teachers' self report. They were questioned on the organization of their classroom and asked how they believed children learned, in terms of two dichotomous statements, below,

'Pupils learn mainly by receiving mathematical knowledge mainly from the teacher textbook and other sources' or

'pupils play an active role in mathematics learning constructing their own knowledge'

(Newstead, 1998, p.60). A greater proportion of pupil led interaction also defined an alternative teaching style as opposed to the traditional style defined by a greater proportion of teacher led interaction.

Newstead investigated the relationship between children's attitude to mathematics and the type of teaching in the class. She found that one of the items eliciting the highest anxiety was the teacher asking the pupils questions and further analysis indicated a social or public aspect of mathematics anxiety, which she called Social Mathematics Anxiety – the answering and questioning in the group in front of peers and teachers. However it does not appear to have been recognised by researchers such as Chiu and Henry who revised and validated a Mathematics Anxiety Scale for Children (MASC). In a large scale study carried out in 1990, Chiu and Henry used the scale with 562 children, and found that mathematics anxiety was greatest for those with learning difficulties and least for an able group (there is little detail about how the children were selected and it should be noted that these groups, totalling 160, were selected from the larger sample.) However, no item in the MASC appeared to focus on the element of questioning in the whole class situation; there was no item directed at questions and answers in the whole class context in the MASC.
Newstead’s sample of 246 children between 9 and 11 years of age found that there was a greater degree of Social Mathematics Anxiety in the more traditional classrooms where a greater number of factual and computation closed questions were asked. Social Mathematics Anxiety appeared less in the alternative classrooms- those with fewer factual and recall questions, and more questions to do with concepts and problem solving. However, there appear to be several areas of Newstead’s study worth questioning. There are methodological issues in the division of teachers into the traditional and alternative groups. Of the 8 teachers 2 were excluded, one because her self report was not in line with her observed practice, and the other because there was not enough pupil initiated interaction. Newstead admitted that the observer judgement was subjective and based on a single lesson observation.

Secondly, it is possible that this Social Anxiety did not appear in the alternative classrooms because there was less exposure of the children with difficulties. In these classes there was a greater proportion of pupil led interaction than in the traditional classes. It is possible that the pupil led interactions did not come from the more anxious children; in that case those children most prone to mathematics anxiety would have had a lower level of interaction in the alternative classrooms with more pupil led interaction, than in the traditional teacher led classes.

Thirdly, Newstead did not measure pupil learning in any way. She acknowledged the failure to gain access to performance data as a flaw in the study and suggested that the relationship between mathematics anxiety and performance needed clarification. She did not speculate on the possible interdependence of the two variables - that while the less traditional classrooms may have less mathematics anxiety there may also be less pupil learning, that alternative classrooms may be related to less learning. Saxe et al., (1999) found that in what they termed reform classes, similar to
Newstead’s alternative classes, where problem solving approaches were used more frequently than in traditional classes, pupils’ performance on problem solving was better.

However, despite these reservations there is no doubt that Newstead’s concept of Social Mathematics Anxiety is an interesting and relevant issue in mathematics questioning in a whole class setting: ‘some pupils were anxious only about the social public aspects of doing mathematics’ (Newstead, 1998, p.66). In common with Butterworth (Butterworth, 2002), who considered the view of the child in the numeracy lesson through focus groups and with Boaler (1997) through a series of interviews with pupils, Newstead has taken account of the child’s view of mathematics. It may be speculated further that the numeracy session, starting daily with a formal period of teacher questioning in a whole class context, may add to the element of Social Mathematics Anxiety identified by Newstead.

These three studies have made implicit or explicit assumptions about higher order thinking in problem solving or conceptual analysis, and computation which seems to be regarded as lower order thinking (Galton et al., 1998, Perry et al., 1993 and Newstead, 1998). Nowhere is higher order thinking defined, although it may include for both Boaler (1997) and Newstead (1998) a metacognitive ability to explain operation choice or strategies. Returning briefly to Table 1.3 we can see it defined as application of rules, the ability to explain, and hypothesis formation, and conceptual knowledge (Boaler, 1997; DfEE, 1999; Kerslake, 1982; Perry et al., 1993).

In contrast to this construct of ‘higher order’ thinking, believed to be more complex than computation, Dowker (1995) analysed the competencies of three children referred to her for difficulties in arithmetic – computation. These children both showed what Perry et al., (1993) would identify as conceptual thinking or higher order
thinking, but neither could carry out computation. One example was Laura, 6 years and 3 months. She could use the commutativity principle, \( n+1 \) and \( n-1 \) to derive answers. [Commutativity is knowing that if \( 9+4=13 \), \( 4+9 = 13 \). \( n +1 \) means knowing that is \( 24+23=67 \) then \( 23+25=68 \); \( n-1 \) means that if \( 9+8=17 \) then \( 9+7=17-1 \) or \( 16 \).] Her estimation was good for single addends and reasonable for double addends. She could not explain her use of commutativity or \( n=1 \) and \( n-1 \) beyond saying 'it's quite close'. Her arithmetical abilities for numbers adding up to 6 or more were very poor. Another child, Paul also 6 years and 3 months could estimate the sums of 2 digit addends. He too could use the \( n+1 \) principle and explain it but he could not add up to more than 10, nor could he arrange counters to show how \( 3 \) add \( 2 = 5 \). Laura could estimate but could not explain, Paul could explain the \( n+1 \) principle but could not carry out small addition sums.

These two cases indicate that it may be misleading to claim that certain types of questions or mathematical process are ‘higher’ order or ‘lower’ order. It is worth speculating on the part played by language ability in what these researchers have termed the ‘higher order ‘ constructs. It may be that different processes are used in addition and subtraction from those used in problem solving and if so, it appears more useful, as Ashcraft does, to consider the processes in reaching an arithmetical answer (Ashcraft, 1992), and the part played by language (Huttenlocher et al., 1994; Snorre, 2002).

1.5 Conclusions

Research evidence discussed above (e.g. Shalev et al., 1998, Butterworth, 2002) Snorre, 2002; Huttenlocher et al., 1994) pointed to the numbers of children who experience persistent difficulties in numeracy. Evidence of additional difficulty with
story problems indicated the part played by language processing and familiarity (Snorre, 2002; Huttenlocher et al., 1994). It appears probable that a proportion of children experience difficulties in the whole class questioning in the numeracy session.

Within the literature reviewed above, there is little evidence of consensus having been reached with regard to the classification of question types. There are assumptions about certain questions producing 'higher order thinking' (Galton et al., 1999, Perry et al., 1993; Newstead, 1998). These questions are in some way superior to questions requiring recall or calculation. This assumption does not appear to be based on psychological theory.

The first research aim of this study was to derive a question classification system for the questions used in the numeracy lessons and the Pilot Study set out to do this. The main study set out to investigate the types of question being asked in the numeracy session and the relationship between these questions and learning. The research aims were to consider whether there would be systematic variation in question type, i.e. teaching style (Newstead, 1998) and how this might relate to learning; the data will be examined to see whether there are systematic variations in questioning. A further research aim is to consider the relationship between distribution of question types and children's learning, examined through a post-test performance.

From the research aim of considering whether questions in mathematical lessons could be categorised and linked to models of mathematical learning the following hypotheses were developed:
firstly, that in keeping with the ORACLE studies (Galton et al., 1998) and those of Perry et al., (1993), teachers would ask more computation questions than problem solving questions. Such computation questions relate to the emphasis on practice in the Numeracy Strategy and on the role of automaticity in arithmetical processing (Ashcraft, 1992; Butterworth, 1999). The hypothesis is derived from the research aim of considering how question types may relate to children's learning in Numeracy.

secondly, that in keeping with estimates (Butterworth, 2002; Shalev et al., 1998; Snorre, 2002) between 2% and 10% would not have been able to access the learning in the whole class session. A test given to children following the numeracy session would reflect this difficulty. This hypothesis is derived from the research aim of considering what proportion of the class may have difficulty with the numerical concepts presented to the whole class.

thirdly, that a larger group of children would have difficulty with the problem solving questions than with the computation questions (Huttenlocher et al., 1994; Snorre, 2002). Errors would be examined in the post-test; in keeping with Siegler's (1987, 1988) and Ashcraft's (1992) findings, errors may give some sort of indication of understanding of the questions and of the strategies and processing used to answer them. This hypothesis is derived from the research aim of considering whether language plays a part in the additional difficulty some children may experience. Error analysis relates to the research aim of discovering what strategies children may be using to answer numeracy questions.
Chapter 2

Pilot Study: Classifying teachers’ questions in the numeracy lesson

2.0 Abstract

The aim of the Pilot Study was to derive a reliable and comprehensive classification system which could be applied to teachers’ questions in the whole class teaching part of the numeracy session. Questions asked by teachers during the first part of the lesson were transcribed and classified. A question classification was drawn up, using categories from the studies of mathematical question types discussed in Chapter 1, the Literature Review. The researcher and three independent judges categorised the questions into four types. There was a high level of concordance between judges indicating reliability, but difficulties with establishing the validity of question classification systems are discussed. The distribution of question types appeared to be in keeping with other studies; additionally there appeared to be an inequality in the proportion of questions directed to boys and to girls in part of the sample.

2.1 Introduction

2.1.1 Question classification

Various classification systems were discussed in Chapter 1, and are summarised in Tables 1.3 and 1.4. Within the literature reviewed above, while there was consensus on some types of questions, there were methodological issues about how the classifications were derived, and their reliability. Most studies offered definitions of the different types of questions but did not explain how the types were derived, nor did they attempt to measure agreement between judges on question typing (Askew
There were confusing differences in terminology which masked similarity of types, particularly apparent in the problem solving category. However, three types of question relating to numeracy emerged from the classifications of mathematical questions (see Table 1.3 in Chapter 1, Section 4). A fourth type relating to class management emerged from Kerslake’s classification questions (Kerslake, 1982) and in more general question types, (see section 4, Table 1.4 in Chapter 1) (Brown and Wragg, 1993; Galton et al., 1999; Mills, 1996). The categories below are derived from these studies; the first three relate to numeracy and the last to class management (see Tables 1.3 and 1.4):

- **factual or recall of facts**

  This category of mathematical question type is used by Askew and Wiliam, 1995; Boaler, 1997; Kerslake, 1982; Ofsted, 2000; Perry et al., 1993. It is also used in the classification of general questions, that is, not specifically mathematical in Galton et al., 1999; Kyriacou, 1995; Mills, 1996; it also appears in Morgan and Saxton, 1994 although they use the less precise term of eliciting information to describe this category.

- **Computation**

  This category of mathematical question type is used by Askew and Wiliam, 1995 using the term calculation, by Kerslake, 1982, and by Perry et al., 1993, although they have combined this with rule recall. It was decided to assume computation rather than retrieval for calculations in view of Ashcraft’s model (Ashcraft, 1992) Chapter 1. Section 2, and Table 1.1 showing research by Siegler, (1987a) indicating that the majority of children of 7 were not using retrieval.

- **problem solving**

  This category is used widely in both mathematical question typing (Table 1.3) and general question classification (Table 1.4). For use with mathematical questioning it includes questions requiring reasoning (Askew and Wiliam, 1995) application of rules
requiring reasoning (Boaler, 1997), word problems (DfEE, 1999; Snorre, 2002) application of facts, hypothesis formation, applied reasoning and enquiry, for example of data, (Kerslake, 1982). It includes computation in context, (that is, word problems), problem solving strategies and conceptual knowledge (Perry et al., 1993) and process questions (Newstead, 1998). It is used to apply to questions which require the selection of the correct mathematical operation and application of knowledge (Ofsted, 2000). When applied to more general questioning, it is the open question classification of Galton et al., (1999).

- class management

This category of question is used to apply to task supervision and routine management (Galton et al., 1999), is termed class management (Kyraciou, 1995), the more general term of management (Mills, 1996). It is defined as the type of question used to establish and control groups (Morgan and Saxton, 1994).

Examples of the categories follow.

Questions which require recall of facts require no working out of mathematical concepts: e.g. 'What times table have we been working on? What is the name of this shape?' These feature in both the classification of mathematical questions in Table 1.3 and the more general questions in Table 1.4. They are closed questions with only one answer.

Questions of computation require the pupil to work out or calculate an answer using numerosity, knowledge of numbers and their relationship to each other, e.g. How many threes in 18? What are the factors of 10? These feature in the classifications of mathematical questions in Table 1.3 and are closed questions with only one
Questions requiring problem solving require the application of mathematical knowledge to produce a new ending or draw a new conclusion e.g. 'What do you notice about the patterns of numbers here? What questions could you ask of these data?' Problem solving questions may be closed or open, with a range of solutions, and their importance was emphasised by Ofsted in an early review of the Numeracy Strategy (Ofsted, 2000). They are perceived to be more challenging, requiring a student to select the correct mathematical operation, hypothesising, and the use of conceptual knowledge and are characterised by a range of terminology in Table 1.3. The category includes selection of the correct operation (Ofsted, 2000), application of facts and applied reasoning (Kerslake, 1982), and the application of rules requiring reasoning (Boaler, 1997). The difficulty which may arise with varying terminology can be seen most clearly in the term 'computing in context' (Perry et al, 1993, p35). As this requires the child to choose the correct operation, it is included in the problem solving category, rather than in the computation category. Examples of problem solving questions are: 'Biscuits come in packs of 6. A carton has 24 packs. How many biscuits are there in a carton?' (Askew and Wiliam, 1995, p.16–17) 'There are 20 teams in a football league. If every team has to play every other team how many games are there in total?' (DfEE, 2000, p.9, Guidance on Teaching Able Children).

Questions of class management are often rhetorical or action-related, focusing primarily on managing children's behaviour or attention. These are designed to facilitate the task, (Brown and Wragg 1993), and were defined by Mills (1996) as questions designed to, 'control pupils' behaviour or direct pupils' attention to teacher or task' (Mills, 1996, p. 5), and this definition was adopted for the category of question called class management. Examples are, 'What about looking a line up
2.1.2 Reliability and validity

The reliability of this classification system, that is whether it is consistent over time, or when used by different people, (Miller, 1984) may be established by using a number of judges or judging the classification system over time, i.e. leaving an interval between classification of questions. It was decided to measure this by using agreement on classification among judges.

Construct validity refers to the correspondence between the operational measures and the theoretical construct which is being measured (Miller, 1984) and is more difficult to establish. Although different types of validity are described (see Gross, 1992) it has been argued that validity is actually a unitary construct, (Airasian et al., 1999,) the central question being whether something measures what it is supposed to measure – the theoretical construct. The theoretical construct in this case is the type of cognitive processing demanded by the different question types, and that is unknown unless each child is questioned about how they arrived at the answer, or through error analysis (Ashcraft, 1992). It is acknowledged that construct validity appears to rest on the question classification adequately and comprehensively describing all of the types of questions in the numeracy session but that there is no independent criterion against which to measure the classification system. To some extent validity is a problematic issue for any classification system; for example the validity of the classification of mental illness, and indeed the lack of agreement on the classification of certain physiological illness (Gross, 1992; Taylor and Hayes, 1990). The issue of validity has similarly dogged the testing of intelligence (Richardson, 1990). The lack of an independently derived criterion is a difficulty for any
classification system where there is no observable or quantifiable difference in the items classified. However, it may be possible to create a construct to validate the system by asking teachers what the intention was in asking the question, and by asking the child what thinking was done to reach the answer. The issue of validity is further considered in Chapter 2, Section 4.3. and in Chapter 6, Section 2.

In conclusion, the aim of the Pilot Study was to classify teacher questions used in the teaching part of the numeracy session to provide a reliable and comprehensive categorisation system.

2.2 Method

2.2.1 Design

Teachers' questions during the first part of the numeracy session, devoted to whole class teaching, were transcribed and classified. Five mathematics lessons in Key Stage 2 (years 3 and 4) were observed, transcribed and the questions coded by the researcher and three independent judges to assess whether the question types were reliable and comprehensive. It was felt that three judges, rather than one other besides the researcher, would be more likely to reveal sources of disagreement in categorising or dissatisfaction with the categories.

2.2.2 Participants

There were three groups of participants, the teachers and pupils, and the judges. The teachers and children were from five year 4 classes in first schools. Four of the teachers were female and one was male. The ages of the children were between 8 years, 7 months and 9 years, 6 months, in the two classes without vertical grouping (without children of year 3). Where there was vertical grouping the ages of the children were between 7 years 6 months and 9 years 5 months. Class size was
between 22 and 28. The schools were in a city area and the catchment area was mainly comprised of local authority housing.

The judges were four adults between the ages of 20 and 52, two male and two female. All were graduates, and two of the four had studied mathematics in Higher Education. The decision was made to avoid using teachers as judges because of emphasis on questioning in teacher training and in recent government publications (for example, The Numeracy Strategy, 1999).

2.2.3 Procedure

Permission was gained from head teachers and staff to be observed. In keeping with the principles of informed consent, it was explained that the researcher was interested in questioning during the numeracy session. If teachers felt uncomfortable at any point the observer would withdraw. The transcript of the lesson was available for teachers if they wanted it. A letter informing parents and requesting consent was sent out from the school and a contact number given for the researcher. The researcher was known to the staff and some of the pupils. She observed the whole class teaching part of the numeracy lesson and transcribed all questions. In each case the numeracy session took place in the morning; the whole class questioning session was the first part of the lesson, and lasted between 20 and 30 minutes.

2.2.4 Classification of questions

A question was defined both by its grammatical structure, i.e. the verb preceding the subject of the sentence, and as a statement requiring an answer. (Galton et al., 1999). Most were in question form, but occasionally the grammatical form was a statement or command. Some questions were rhetorical - a management device rather than a genuine question requiring an answer from the pupil, - and these were
classified as questions if the grammatical form was that of a question. A description of the question types was given to the judges with examples:

- factual or recall of facts: questions which require no working out of mathematical concepts: e.g. 'What is the name of that shape? What times table have we been working on?'
- numerical calculation/computation or working out: questions asking the pupil to use a knowledge of number to answer a question – usually requiring calculation or computation e.g. 'How many threes in 18? What are the factors of 10? What is 10 less than that?'
- problem solving: questions requiring the application of mathematical knowledge to produce a new finding or draw a new conclusion e.g. ‘What do you notice about the patterns of numbers here? What questions could you ask of these data?'
- class management: questions which are often rhetorical or action-related focusing primarily on managing children’s behaviour or attention, e.g. ‘What about looking a line up (using the number lines)? Shall we use our whiteboards? and discipline questions e.g. ‘Would you be happier sitting here, Rebecca? Shall we wait for Daniel to stop talking?’

2.2.5 Independent categorisation of questions

The judges were asked to code the transcribed questions using the above categories or to leave a question mark if they could not easily assign a question to one of the categories. The task took between 45 minutes and 75 minutes. The judges carried out their classification individually. Judgements were analysed for agreement. Following their classification, they were asked:

Did you find you could allocate all the questions to a category?
Did any other types occur to you? If so, what?

Can you tell me about any questions which were difficult to classify?

Questions in which there were discordant ratings were discussed with the judges.

2.3 Results

2.3.1 Inter-judge agreement in classification of question types

Judges’ categorisation of questions yielded nominal data, requiring a non parametric test of agreement among the four judges. To measure the concordance of rating on each question it was necessary to carry out Cohen’s Kappa, (Siegel and Castellan, 1988) which analyses the level of agreement question by question, rather than overall agreement on numbers allocated to each question type. Cohen’s Kappa measured the agreement for all questions, yielding a K value of 0.781 (N,133). Perfect agreement would have been 1, and no agreement would have been 0. Converted to a z score this gave $z = 30.682$, $p < 0.0001$.

It can therefore be concluded that there was a significant level of agreement among judges.
Figure 2.1 Question classification by judges, showing the pattern of judges' classification of questions.

2.3.2 Concordance and areas of lack of concordance between judges

On the vast majority of questions there was agreement. On 29 questions of 133 there was disagreement, but this was from a total of over 532 ratings, 4 per question. On 27 of the 29 questions there was one judge of the four who differed, and on two questions there was an even split between the judges. The main areas of disagreement were that judge 1 classified more questions as factual than problem solving, and that he allocated repetition of the same question to class management; he assumed that when a teacher called out several names in succession this was to focus the attention of those children. In fact, it followed a 'don't know' or wrong answer and was the same numeracy question repeated. This was resolved in later discussion. Examples of questions which judge 1 classified as factual and the other judges classified as problem solving were:

'How did you work that out? Can you explain that to the rest of the class? How did you do that? Did you do that in your head?'
2.3.3 Questions of factual recall or questions of computation

While there was no obvious lack of concordance in other areas, one judge raised the issue of allocating a question to the factual recall or calculation categories (see discussion). The difficulty arose from deciding on the process children were using to answer questions, recall of a learned fact, or computation. The example raised by one judge was the question: 'What is 3 times 2?' He judged this to be recall of a learned number fact, part of a table, while others judged it to be computation, working out two lots of three. Two other such questions were: 'What is one less than 13? What are the factors of 6?'

2.3.4 The reliability and validity of question types

As the concordance of agreement of judges was highly significant it may be assumed that the classification system was reliable. Furthermore it served to classify all questions; no judges left question marks and all said that the questions could be matched to one of the four categories. However, the issue raised in the preceding paragraph indicates that the classification of questions depended on what mental operations the judges believed the children would have to do to answer the question. This issue that of not knowing what mental processes were used to answer a question, will be discussed further in Chapter 2. Section 4.3 and in Chapter 6, Section 2 on validity.

2.3.5 Question Types

It can be seen from Table 2.1 and Figure 2.2. that computation or calculation questions were most frequently asked, followed by problem solving; there were
similar proportions of factual or recall questions and class management questions. Computation questions were approximately double any other category.

<table>
<thead>
<tr>
<th>Fact/Recall</th>
<th>Computation</th>
<th>Problem solving</th>
<th>Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>24.3</td>
<td>56.5</td>
<td>28.0</td>
<td>24.3</td>
</tr>
</tbody>
</table>

Table 2.1  Mean ratings for each type of question

Figure 2.2  Percentages of question type

2.3.6  Gender differences: mean numbers of question types asked to boys and girls.

Although this had not been the intention of the pilot study it was noted that one teacher in the sample appeared to ask questions of more boys than girls, and in the remaining two lessons observed for the pilot study, the researcher noted whether questions were asked of boys and girls. Means differed for boys and girls in all question types but it was thought inappropriate to assume this data was representative for the following reasons. Firstly, the majority of questions in this
sample could not be assigned to either a boy or girl as many questions were directed at the whole class, or the class was encouraged to answer in a chorus, or they were directed at several pupils in quick succession. Secondly, the sample of lessons observed in which the gender of the child was noted was very small (two). However, inspection of means indicated that in this case, for each category of question boys were asked more questions than girls (Table 2.2).

<table>
<thead>
<tr>
<th></th>
<th>Factual</th>
<th>Computation</th>
<th>Problem Solving</th>
<th>Class management</th>
</tr>
</thead>
<tbody>
<tr>
<td>boys</td>
<td>3.4</td>
<td>9.4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>girls</td>
<td>2</td>
<td>6.6</td>
<td>1.8</td>
<td>2.8</td>
</tr>
</tbody>
</table>

Table 2.2 Mean question types to boys and girls

2.4 Discussion

2.4.1 Classification of questions

There was a significant level of agreement among judges on allocation of questions to categories. Of 532 ratings, only 5.82% were discordant. One judge was responsible for the majority of the lack of agreement and these differences were resolved in discussion later. He had allocated more questions to factual than the other judges and repetition of questions to class management – i.e. where a child could not answer, the teacher would ask several children the same question in succession. The other judges allocated such repetition to computation, or problem solving – whichever category the original question had belonged to. In later discussion judge 1 said he had not realised that they were repetitions of a question but thought the teacher was calling out names to redirect the child’s attention. Had the judges been able to watch a video of the session such a misunderstanding would
not have arisen. The methodological implications are, therefore, that judges require a transcript and either a video tape recording, or an audio tape recording to make the context of the question clear.

2.4.2 Videorecording as an observational method

However, there are issues which arise from video recordings. There are two clear advantages for a study of questions such as this; the tape can be played and replayed to gain the most accurate transcript of teachers' questions, and it can be judged by an independent observer. For classroom research the opportunity to view and review an interaction is of great value, despite the drawback of the additional time:

'\textit{the video does offer a relatively unfiltered record of transactions -- and a permanent detailed record is provided}' (Simpson and Tuson, 1995, p.52) although the point has been made that filtration happens with the location and timing of the recording (Simco and Warins, 1997).

The difficulty of trying to achieve objectivity has been discussed in the exploration of validity in image based research (Simco and Warins, 1997). The authors have questioned whether the findings were genuine for the group represented and discuss two video clips of classrooms with alternative interpretations of what is happening. It is acknowledged that the presence of a videotape is likely to affect the behaviour of both teacher and pupils, but it may be argued that the presence of an observer does this also. The possible costs and benefits are discussed further in Chapter 6.

2.4.3 The validity of the classification system

An issue which both mirrored the debate about 'drill and practice' (Butterworth, 1999; Keys et al., 1997) and understanding, and touched on the question of validity, was
raised in discussion afterwards by judge 1, who felt that certain questions required recall of a learned number fact rather than calculation. In discussion with the judge following the rating exercise, this appeared to stem from the fact that he believed that times tables and simple addition/subtraction facts would be recall or fact rather than 'working out' that is, numerical calculation. The first issue here is the underlying belief about the basis of numeracy; the argument centres around the importance of memorising number facts - bonds and tables, and the understanding underlying such facts. This is discussed by Butterworth (1999) and in the Preliminary Report of the Task Force (DfEE, 1998), and will be explored further in Chapter 6.

The second issue raised by this debate is the validity of any classification system which reflects the assumed cognitive processing carried out to arrive at an answer. For example, in one lesson a child was asked whether a number was odd or even. The child may ask herself, ‘Does it divide by two?’ or may know by looking at the number whether it is odd or even. Sometimes it may be evident what the child has done from the speed of the answer. For example, a boy answered quickly 10001 when asked for an odd number, and so the question was defined as requiring recall of fact. However, it is acknowledged that the classification makes assumptions about the cognitive processing to reach the answer. To some extent this is a factor in many psychological experiments; information gained from error analysis yields indirect information on how the child has reached the answer (Ashcraft, 1992) and so it was decided to scrutinise errors in the main study. However, the issue of validity of the classification system raised here and above in Chapter 1, Section 4 is considered further in Chapter 6, Section 2.
2.4.4 Question type

Although the purpose of the pilot study was to establish a classification system, it is of interest to consider briefly the pattern of question types emerging in this sample. Inspection of the means showed that the largest category was computation or calculation questions. Recall of fact questions were fewer, but tended to be in similar quantities to class management questions.

2.4.4.1 Recall of fact questions

The difficulties inherent in allocating certain questions to the factual or recall category are discussed above, and are at the root of some of the discussion of the role of rehearsal and memory in children’s Numeracy (Keys et al., 1997, Butterworth, 1999). Speed of response may eventually depend on ‘knowing by heart’ rather than ‘figuring out’ but with children of seven and eight it may be reasonable to assume that for most of them such questions involve a calculation rather than recall of a fact. Three of the four judges believed that children of this age would not know numerical facts by heart. It is interesting to note at this point that the confusion also exists for the government which has assumed that questions were aimed at recall rather than computation:

‘Where the teaching was unsatisfactory, too much time was spent on the recall of number facts rather than expecting pupils to figure out new facts from known ones and explain how they had arrived at their answers.’ (DfEE, 2000p.6, section 6)

2.4.4.2 Computation questions

The largest single group of questions required a computation or calculation, such as, ‘How many threes in eighteen?’ and ‘What are the factors of sixteen?’ The fact that this sample of teachers’ questions in the numeracy hour shows the largest number of
questions required a simple calculation is in keeping with the comments made above in an early review of the numeracy strategy (DfEE, 2000), assuming that this comment included questions aimed at calculation. There were approximately twice as many computation questions, as questions in any other category. If we add to this the questions requiring factual recall as the DfEE report (DfEE, 2000) appeared to have done, this amounts to sixty per cent of the total and may partly explain why the report indicated that teachers were not asking enough problem solving questions:

'More importantly, some teachers have not yet got the balance right between asking closed questions which require nothing more than rapid recall and more open questioning which, for example, asks pupils to explain how they would set about solving a problem' (DfEE, 2000, p 9, section 29)

2.4.4.3 Problem solving questions

What defined problem solving questions? Judges' categorizing showed high concordance. Two examples show that children were required to deduce rules from patterns appearing in numbers, and ask questions of data:

'Look at the numbers which make up that sum. What do you notice?'

'What do you notice there? (Answer: There's only one even number)

'So an odd number of odd numbers gives you an odd number?'

'If you were presented with this frequency table what sort of questions could you ask?'

Questions like this ask the children to deduce something from the data they have. In the first case the teacher wanted them to deduce something about the pattern of numbers required to make up an odd number. In the second example the teacher
wanted the children to look at a frequency table they had just drawn up on the whiteboard of favourite foods using a tally chart. The problem solving question was designed to persuade pupils to scrutinize the information they had gained to see what they could ask of it. During observation it was noticeable that some teachers tended to give such questions a different status from the numerical calculations prefacing them by remarks which appeared to be designed to alert the children. Four teachers prefaced them by remarks such as:

'Here's something I really want you to think about,' and 'Now, think carefully all of you' or 'Just imagine you had this table — all this information in front of you— I want you to really think'

Even without this additional contextual information judges appeared to have little difficulty in deciding which questions were problem solving and no issues were raised about this category in discussion.

2.4.4.4 Class management questions

This category accounted for approximately one quarter of all questions. Examples of such questions were:

'Shall we all use our whiteboards now? Can you all have a whisper to your partner to decide how to tackle this? Would you be better sitting nearer to me, N?' Are you all ready to think about this question now?

It may be that this group of questions could be further subdivided into questions to facilitate the carrying out of tasks and questions to direct the attention of specific children to the task. The definition used by Mills (1996), to: 'control pupils' behaviour or direct pupils' attention to teacher or task' (Mills, 1996, p. 5) contained these two elements. However, it is worth noting that there were considerable teacher variations
in the use of class management questions. In the case of one teacher these questions made up nearly 30% of the total number of questions, while for another they accounted for less than than 20%. It may be that class management is carried out in different ways; for example teachers may tend to use statements as vehicles for class management or numeracy questions directed towards those children the teacher perceived as being off task.

2.4.5 Teacher variation and possible gender differences

This was a small sample with considerable teacher variation. One teacher asked very few questions but made a much greater proportion of statements; one teacher asked a greater proportion of class management questions but fewer questions in other categories. However, it did appear that one teacher directed more questions towards boys than to girls. For the reasons given in Chapter 2, Section 3.6, above, no conclusions can be drawn. Gender had not been included as a factor in the original design, the sample was too small, and incomplete. The means indicated differences which were most marked in the problem solving category, but the sample was of only two lessons.

Certain studies have showed that teachers may question boys and girls differently; for example Barba and Cardinale (1991) and Webb (1984) found that in groups of four males, pupils received more explanations from peers and the teacher, than did females and that females were more likely than males not to get answers. Leder (1990) commented on various studies in the 1980s in Sweden, American, Australian and English classes showing a bias to interactions with males but noted that the differences were small and few were significant. She carried out a study in Australia in 1990, using a video camera and sampling lessons from grades 3, 6, 7 and 10, and requested the teachers include as much oral work as possible. She found boys at
each level interacted more with the teachers. At Grade 3, around age 8/9, significantly more low cognitive level questions were directed at boys and more discipline exchanges. Overall, there were more exchanges with boys than girls although this was not significant for high cognitive level questions. The differences in interactions with the teachers at each age were in favour of boys and were most significant at grade 10. However, she has adopted a low and high cognitive question categorisation, and it is not clear how these were coded and by whom. This raises again the issues considered in Chapter 2, where questions have been assumed to tap high or lower order cognitive processing.

Whatever the difference in interaction with boys and girls, there appears to be little evidence for a significant gender gap in performance. A review of 98 studies of differences in mathematics achievement showed that the average difference is small and decreasing (Friedman, 1989 p.32): though there were small differences, these were mostly among higher attainers. More recently Gorard (2001) analysed data from Wales and indicated the gaps were small and mostly at the high attainment end. Leder (1990) too found no difference in performance until Grade 10 when boys did better than girls. Mathematics appears to be more important to boys and to be connected with male career aspirations (Campbell and Evans, 1997).

It is appreciated that other factors may have an effect on teachers’ questioning of boys and girls. For example it may be that girls are more reluctant to join in whole class teaching interactions. Boaler (1997) has drawn attention to the dissatisfaction of girls with traditional mathematics teaching methods and it may be that in the relatively formal and public arena of the numeracy lesson girls are less likely to volunteer answers. Although no firm conclusions could be drawn from this small sample, in the light of the differences in mean numbers of questions directed to boys and girls, it was decided to investigate possible gender differences in the main study.
2.5 **Limitations and conclusions**

The pilot study had certain limitations which influenced the design of the main study. Firstly, coding by judge 1 showed that fuller knowledge of context was sometimes necessary to categorise questions. However, in cases of disagreement, when the context was described to judges in detail, they were able to reach consensus. Secondly, the issue of gender, that is gender of child questioned, was raised towards the end of the observations. The remaining two lessons were too small a sample to yield any firm conclusions. Thirdly, the practical difficulties of assigning teachers' questions to boys or girls were considerable. For example, the questions were often directed to the whole class or a group (one table for example). If a boy or girl repeatedly volunteered the answer, and this was often the case, it cannot be concluded that the teacher meant to question that person, merely that the teacher selected the child whose hand was up. Equally, the teacher may have selected a boy first who did not know the answer, and subsequently asked a girl with her hand up. Assignation of questions was to boy or girl when this was clear, whether or not it was for the first time of asking, but it was decided that questions directed at the whole class or to a group could not be assigned fairly to either gender; it could not be claimed that the teacher intended to direct the question to a boy or girl but rather was constrained in choice by which children actively volunteered the answer. In the light of such limitations it was decided that although intrusive it was necessary to record the lessons on videotape so that the context would be clear.

In conclusion, there was a significant level of agreement among the four judges on allocation of questions to the four categories, and where judges had disagreed this was resolved in discussion. This categorisation could be usefully and reliably used to describe questions in the numeracy lesson. Means for the sample indicate that the
single largest category of questions required a straightforward calculation or computation. This categorisation system would be used to investigate the types of question being asked in the numeracy session and the relationship between these questions and learning (see Chapter 1, Section 5).

The original research aims were to consider the types of questioning used in the whole class teaching in the numeracy session and to endeavour to relate these to the learning of the children. A further research aim which developed from the literature review was to explore the learning of the children and the proportion who may be experiencing difficulties in understanding the concepts taught in this first part of the lesson, often through questioning. The hypotheses derived from these aims were:

- firstly, that in keeping with the ORACLE studies (Galton et al., 1998) and those of Perry et al., (1993), teachers would ask more computation questions than problem solving questions. This was supported by the pilot study.

- secondly, that in keeping with estimates (Butterworth, 2002; Shalev et al 1998; Snorre, 2002) between 2% and 10% would not have been able to access the learning in the whole class session. A test given to children following the numeracy session would reflect this difficulty.

- thirdly, that a larger group of children would have difficulty with the problem solving questions than with the computation questions because of the language used (Huttenlocher et al., 1994; Snorre, 2002). Errors would be examined in the post-test; in keeping with Siegler’s (1987, 1988) and Ashcraft’s (1992) findings, errors may give some sort of indication of understanding of the questions and of the strategies and processing used to answer them.

- In addition to the hypotheses outlined in Chapter 1, Section 5 a further research question was the relationship between children’s performance in
post-test questions and the teachers' use of different types of questions (Saxe et al., 1999).

- A further research question arising from the pilot study concerned gender and the direction of teachers' questions. While some research evidence, discussed in Chapter 2, Section 4.5 has indicated teachers may question boys and girls differently (Barba and Cardinale, 1991; Webb, 1984) evidence from the pilot study was limited, so it was decided to analyse the data in the main study for direction of question and gender. Research by Gorard et al., (2001) has also indicated no evidence of difference in gender attainment and it seemed probable that this would be reflected in post-test performance.
Chapter 3

Method

3.0 The Research Design

The research aims were to analyse the types of questioning used in the whole class teaching in the numeracy session and to endeavour to relate these to the learning of the children. The classification system derived in the pilot study was used. A further research aim which developed from the literature review was to explore the learning of the children and the proportion experiencing difficulties in understanding the concepts taught in this first part of the lesson. The hypotheses derived from these aims and from the literature review are summarised at the close of Chapter 2, Section 5.

The design was the observation and classification of questions in ten lessons and the testing of children just after the teaching. The distribution of teacher question types and the children's tests results was analysed to see whether there was a relationship between question types and learning. Errors in test results were further explored for clues to the strategies children were using.

3.1 Participants

Participants were ten teachers, eight of whom had taught for four years or more, one was in her second year of teaching and one in her third year. There were eight females and two males, between the ages of twenty-six and fifty-five. Initially an effort was made by the researcher to find classes with nearly equal or nearly equal numbers of boys and girls but this proved impossible. Teachers were approached who were not in schools which were under special measures, following inspection.
One school however, had serious weaknesses, and an action plan following inspection. Where an initial request was made and there was reluctance on the part either of the head or the teachers involved the request was not followed up. For example, in one school the acting head teacher was uneasy about giving permission and although she did not refuse outright the request was not pursued. In nine of the ten schools the researcher was known to the staff through her role as educational psychologist to the school, in the tenth school the request was made through a colleague and followed with an introductory meeting with the head teacher. Teachers of years 4 and 5 were chosen as the interim review of the Numeracy Strategy in 2000 (Ofsted, 2000) noted that results in year 4 children were disappointing. It was felt that this age group should be studied for that reason, and for the additional practical reason that children in years 4 and 5 do not have compulsory standard attainment tasks, which may be a factor in teaching during year 2 and 6. In general school and class routines are better established than they may be in younger classes. There was vertical grouping in four of the schools so that the classes included year 3 children.

3.2 Procedure

Nine of the ten schools were in or served large local authority housing estates on the edge of a city. Their catchment areas were considered to be predominantly from areas of social deprivation. The tenth was a school also on the edge of the same city but serving an area of owner occupied houses, privately rented and local authority housing. The recordings were carried out in the spring and summer terms, 2002. A pilot study to draw up a classification of questions had been carried out during the summer term of the year 2000.

Head teachers of schools were approached by the researcher and given an outline of the research. They were asked whether the researcher could approach the teachers
of years 4 and 5 to request their participation in the research. In all cases head teachers and class teachers asked what the aim of the research was. In each case the researcher told them it was concerned with questioning in the teaching part of the numeracy session, although not that it was about different question types nor that gender was of interest. In line with ethical guidelines, (BPS, 2000) teachers were told that the area of interest was the whole class teaching part of the numeracy session and that transcriptions of the videotapes would be made by the researcher. If teachers wanted to view the tapes they could do so and following that the tapes would be erased. They could also request to see the transcripts of the videos. One teacher asked for the video tape and watched it with her class, and two teachers requested the transcripts.

Following agreement from head teachers and teachers, letters requesting consent were sent to parents whose children were in the classes. No parents refused permission, nor did any contact the heads or the researcher for further information. Dates for the observation had to be set well in advance and on several occasions had to be rearranged because of events outside the control of teacher or researcher (for example, Ofsted visits, class room repair).

A Sony Handycam was used. The researcher positioned herself at the back of the class who were most often on the floor in front of the white board. The recordings lasted for the length of the first part of the Numeracy session, when the teacher was teaching the whole class. The time varied from 14 to 33 minutes. One class sat at their tables. The camera was generally focused on the teacher or on the white board.

Recording tapes were transferred unedited to videotapes and transcribed later. Transcription times were between one and two hours for each tape and was limited to teachers' questions.
The whole class episode was the first part of the numeracy session in which teachers led and supervised the interaction, usually a prelude to the second part of the numeracy session when students worked in groups. Lesson content varied and it was not able to establish what would be taught beforehand in all observations. The areas covered included multiplication of 2 digit numbers, the degrees in a straight line and in right angles, the making of a frequency chart for likes and dislikes of food, a Carroll diagram, and different strategies for adding large numbers.

3.3 Coding of questions

These were coded first by the researcher into four types, derived from the pilot study.

**Factual recall** was a question which could be answered by the child from memory:
‘What was our objective yesterday? What’s an easy way to times by 10?’

**Computation** was a question which needed arithmetical calculation to answer it:
‘7 add what is 10? 3 times 4? 65 add what is 100?’

**Problem solving** was a question which needed the child to decide what mathematical procedure to use, (a problem) or a question which asked the child how s/he would carry out a procedure:
‘How many pencils costing 15 pence can I buy for one pound? How could you take away 345 from 1000 – tell me two ways?’

**Class management** were routine and task supervision questions as well as questions designed to ensure the children were paying attention:
‘Would you be better sitting here? Can you look at the white board? Could you get your white boards?’
3.4 Videorecording as an observational method

The pilot study was undertaken using a tape recorder and observation but, as it was not always clear what mathematical procedure or issue the questions related to and whether a question had been directed to a boy or girl, it was decided to use video recording for the main study. There are advantages and disadvantages associated with the use of video material. One advantage is the opportunity to view and review an interaction (Simpson & Tuson, 1995, p.52) but the difficulty of trying to achieve validity arises in video taped research (Simco & Warins, 1997). In other words, would the ten videotapes be representative of teachers' usual practice in questioning in the numeracy session? This issue is discussed further in Chapter 6, the Evaluation. A further practical issue noted by Simco and Warins is the time taken to transcribe. In this case the time taken to transcribe was generally between 3 to 4 times the length of the recording. For example, fifteen minutes of tape took more than one hour to transcribe and additional time to code. The transcriptions focused on the questions and were not a complete record of all that happened. For clarity the transcriptions included the subject of questioning, what was on the white board, and what Simco and Warins referred to as the general ecology of the classroom (Simco and Warins, 1997). The same authors also pointed out that narrative transcription always contained omission; because of this the researcher watched and transcribed once and then watched again with an independent observer with the transcript. The second viewing did throw up missed questions. Only questions from the teacher were transcribed and analysed. A sample of a transcript is shown in Appendix 1.
3.5 **Assessment of learning**

Immediately after the teaching part of the numeracy session, children were asked to fill in the answers to between 6 and 8 questions referred to as the post-test. These were similar to questions set by the teacher in that session and were of two types, computation and problem solving. Computational questions were often times tables, addition or subtraction, and were based upon the content of the preceding lesson. Because of this post-test questions varied from class to class and the level of difficulty varied according to the content of the lesson (further discussed in the Evaluation, Chapter 6 Section 1). A pre-test would have been desirable to give a measurement of learning in the lesson, but the practical problems which prevented this and the resulting limitations on interpretation of results are further discussed in Chapter 6.

An example of a problem solving question was to ask the children to draw a Carroll diagram based on numbers divisible by 10, and numbers which were not over 100, and allocate numbers to the correct boxes. Similarly they were asked to carry out long multiplication problems using a decomposition method they had just learned, and show how they had done this. The variation in number of questions reflected the length of the lesson. The questions started and finished with a very easy item so that children would not be unduly anxious, nor left with a feeling of failure. Examples of first and last questions were: '4 times 3' (mental mathematics had been the four times table)

and another was '2x7 equals 7x2 True or False'

Children sat in their usual places and were told that they should write down what they thought were the correct answers. They were told that only the researcher would
know what they had written and that they should guess or draw a line if they were not able to answer. They were asked to write their names on their papers, or to write boy or girl, for allocation of the results to boys or girls. Questions were scored with either 1 for correct or 0 for incorrect apart from those which asked children to show how they tackled the question. In that case 1 mark was given for a correct strategy and 1 for a correct answer. Scores were converted to proportion correct, of either computation or problem solving questions.

Jordan (2002) found that untimed addition could be done by most children even with learning difficulties and that timing was a crucial additional factor. Children were given 2 minutes for each question. The questions were re-read after this allowing a further 1 minute for each. The majority of children appeared to finish well within the time.

As the post-test was composed by the researcher during the lesson, a sample of two post-tests were shown to an independent observer, one of the judges for the pilot study, who had been asked to watch the video of the lessons. He was asked to rate how far each of 12 questions reflected the lesson content. Where 1 represented no reflection of the content (not at all) and four represented high correspondence (was very similar) 9 of the 12 questions were rated as 4 (75% had the highest rating possible). One of the first questions was rated as 3 and the final questions in each post-test were rated at 2. The rater pointed out that although the final questions were on the same topic they did not reflect the difficulty level of the lesson questions. This was because for ethical reasons the researcher had ended the post-test with a question that was likely to be answered by all the children.
3.6 Ethical considerations

Ethical guidelines were followed (BPS, 2000). Informed consent was sought from the teachers; this meant that they knew the purpose of the observation and this had implications for how far it was possible to generalize from their behaviour. On one hand it was desirable to recreate as closely as possible the environment of the usual numeracy session. On the other hand, ensuring that participants give informed consent, meant that the teachers were entitled to as full as description as possible: 'it would be quite unethical not to describe as fully as you can the areas of interest of your studies' (Simpson & Tuson 1995, p. 59)

However, although teachers were told that the research was about teacher questioning in the numeracy session, they were not told of the specific investigation of question types, nor of gender. Though participating teachers were offered the chance to view and discuss their own video, only one did so. In one case confidentiality was an issue when the head teacher requested the video. It was necessary to refuse and to discuss the ethics of observation. If either teachers or head teachers appeared reluctant or anxious, no effort was made to persuade them. Video recordings will be erased.

Permission was requested from the parents of the children in the classes; a contact number was given for the researcher for parents to phone for further information. None did so. The children were told after the first part of the numeracy session that the researcher wanted them to answer some questions on paper. They were told that only the researcher would see these answers, and that it they did not know the answer they could draw a line, or guess. The questions started and finished with the easiest questions, to minimize failure for children with mathematical or general learning difficulties. Children had been told that only the researcher would see these
results; this policy was followed. However, four of the ten teachers asked for the questions to go over the answers either in the plenary session that day or the following day. The children were thanked and praised by both the researcher and teacher.

3.7 Question types

Questions were transcribed by the researcher and coded as one of the four types. The teaching part of the numeracy session was transcribed and all questions were categorised into four types – factual, computational or calculation, problem solving, and class management. The post-test data varied in the number of questions set to the classes – either 6 or 8, so the data was converted to proportion correct for each type of question, and the proportion correct for boys and girls separately. Examples of questions are in Appendix 2.
Chapter 4

Results

4.0 Results
This chapter is divided into two parts; part one will deal with teachers' questions during the first part of the numeracy lesson and part two with the assessment of children's learning or post-test.

4.1 Part one: teachers' questions
The data here were a number of different types of questions for the ten teachers. Hypothesis one was that, in keeping with the ORACLE studies (Galton et al., 1998) and those of Perry et al., (1993), teachers would ask more computation questions than problem solving questions. This was supported by the pilot study. The data would be examined for any systematic pattern of variation in question type, i.e. teaching style (Newstead, 1998). Data would also be investigated to see whether there was a difference in the number of questions directed to boys or girls.

4.1.1 Computation versus problem solving questions
The data were different types of questions for the ten teachers, with normal distribution. The Kolmogorov – Smirnov test was not significant: A non significant result, (probability of greater than .05,) indicates normality (Pallant, 2001, p.58) and that the data were suitable for parametric analysis (Howell, 1982; Pallant, 2001). Computation questions, p = 0.44,
Problem solving questions p = 0.25.
To test the hypothesis that teachers would ask more computation questions than problem solving questions, a paired t-test was carried out. It was found that more computation than problem solving questions were asked.

\[ t, (9) = 2.52, p< 0.05 \]

<table>
<thead>
<tr>
<th>Question Type</th>
<th>Mean Percentage</th>
<th>Standard Deviation (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>computation</td>
<td>37.53</td>
<td>11.53</td>
</tr>
<tr>
<td>problem solving</td>
<td>25.81</td>
<td>12.71</td>
</tr>
</tbody>
</table>

Table 4.1  Means for computation and problem solving types of questions

However, the mean figure disguises individual teacher variation; teachers 1,2,3,5,7,8 and 9 asked more computation questions than problem solving; teachers 4 and 6 asked more problem solving questions than computation. The figure below, Figure 4.1, shows the individual teacher patterns.

Figure 4.1  Computation and problem solving questions by teacher, as a percentage of total questions
4.1.2 Distribution of question types asked by teachers: teaching style

To explore whether there was any systematic variation in teaching style shown by distribution of question types, data were examined to discover whether any patterns of questioning could be discerned throughout the sample of teachers. Figure 4.2 shows patterns of question types posed by teachers to children.

![Question Type Percentages for Each Teacher](image)

Note: Percentage, on the Y axis, is the percentage of total questions

**Figure 4.2** Question Type Percentages for Each Teacher

A Cluster Analysis, which picks out similarities and differences in a sample, was performed using hierarchical clustering with average linkage. It revealed a tendency towards two clusters, with one group of seven teachers being differentiated from a smaller group of three teachers. The larger group will be called group A, and tended to ask a greater number of both computational and problem solving questions than...
the smaller group, group B, who asked more factual questions. Figure 4.3 illustrates in the form of a dendrogram the relationship between these groups.

Figure 4.3  Dendrogram of a cluster analysis showing two different groups by use of question type

These two groups will be called questioning style Type A and Type B. Type A is the larger group of 7, and consists of teachers 1, 2, 4, 5, 8, 9, and 10. Type B consists of 3 teachers, 3, 6 and 7. For teachers in group B the largest category of questions was factual; for teachers in group A the largest category was computational, Fig. 4.4.

<table>
<thead>
<tr>
<th></th>
<th>Factual Percentage(%) Means &amp; (SD)</th>
<th>Computation % Means &amp; (SD)</th>
<th>Problem Solving % Means &amp; (SD)</th>
<th>Class management % Means &amp; (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type A</td>
<td>16.91 (8.73)</td>
<td>43.4 (6.58)</td>
<td>30.35 (9.64)</td>
<td>9.29 (6.26)</td>
</tr>
<tr>
<td>Type B</td>
<td>42.8 (10.13)</td>
<td>23.7 (7.88)</td>
<td>15.22 (15.43)</td>
<td>18.21 (7.56)</td>
</tr>
</tbody>
</table>

Table 4.2  Mean percentages and standard deviations (SD) of four question types by teacher groups A and B
Figure 4.4 shows this pattern of outcomes and Table 4.2 shows the difference in mean percentage for each question type, and the two groups, A and B.

![Percentage Means between Question Types](image)

**Figure 4.4** Mean percentages of question categories for type A and type B

The sample size is too small to be able to generalize to a larger teacher population. It is possible that a larger sample might show a similar pattern or that additional patterns might emerge. The possible relationship between learning measured by post-test performance, and questioning distribution is discussed below, in part 2.

### 4.1.3 Relationship between gender of child and teachers’ questions

Research question two, which arose from the pilot study concerned gender and the direction of teachers’ questions. Many of the questions were not specifically directed at either boys or girls but at the class which sometimes answered in unison, or the same question would be asked of a table, or group, or to several children in
succession. Only questions specifically directed to a boy or girl were counted and so formed a smaller subset of data; mean figures are shown below in Table 4.3. The data were percentages of the question type to boys and to girls separately; this was necessary because numbers of boys and girls were unequal. The data were investigated to see whether there was a gender difference in the teacher questioning, i.e. whether there was a difference in the distribution of question types directed at boys and girls. Table 4.3 shows the percentages of each question type directed to boys and girls (it can be seen that only part of the total could be allocated to boy or girl and this is particularly evident in class management questions). Independent sample t tests were carried out on each question type and found there were no significant differences in the percentages of the four types of question directed to boys and girls (Table 4.3).

<table>
<thead>
<tr>
<th>Question Type</th>
<th>Mean percentage</th>
<th>Standard deviation (SD)</th>
<th>T value (t test)</th>
<th>P (probability)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factual boys</td>
<td>21.5</td>
<td>18.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factual girls</td>
<td>25.27</td>
<td>14.34</td>
<td>0.518</td>
<td>0.611</td>
</tr>
<tr>
<td>Computation boys</td>
<td>39.99</td>
<td>20.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computation girls</td>
<td>39.92</td>
<td>18.14</td>
<td>0.008</td>
<td>0.994</td>
</tr>
<tr>
<td>Problem solving boys</td>
<td>30.49</td>
<td>14.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Problem solving girls</td>
<td>23.13</td>
<td>20.69</td>
<td>0.923</td>
<td>0.368</td>
</tr>
<tr>
<td>Class management boys</td>
<td>8.01</td>
<td>11.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class management girls</td>
<td>12.12</td>
<td>18.18</td>
<td>0.610</td>
<td>0.549</td>
</tr>
</tbody>
</table>

Table 4.3  Mean percentages of question types to girls and boys
(i.e. percentage of total directed to girls and to boys)

89
4.2 Part 2: learning shown by analysis of post-test performance

These data relate to the learning of the children during the first whole class part of the numeracy session. This was measured by a post-test with between 6 and 8 questions, depending on the length of the lesson. Data were therefore converted to proportion correct. Hypothesis two was that in keeping with estimates (Butterworth, 2002; Shalev et al., 1998; Snorre, 2002) between 2% and 10% would not have been able to access the learning in the whole class session. A test given to children following the numeracy session would reflect this difficulty. Hypothesis three was that a larger group of children would have difficulty with the problem solving questions than with the computation questions because of the language used (Huttenlocher et al., 1994; Snorre, 2002).

4.2.1 Proportion of children whose performance indicated difficulties.

Hypothesis two predicted between 2% and 10% of the class would show in the post-test that they would not have had significant difficulties understanding the concepts taught in the whole class part of the numeracy session; criteria used were those scoring 25% and below and those scoring 0% in the post-test (see Shalev et al., 1998 for a discussion of these criteria). Graphs in Appendix 3 show the distribution of scores for each class. The number of those scoring 25% or below and zero are shown in Tables 4.3 and 4.4. It can be seen that the 22% of children scoring less than 25% in problem solving is close to the lowest quartile (Shalev et al., 1998) but that the 19% scoring zero in the problem solving questions is considerably more than the lowest 5% deemed dyscalculic by Shalev and colleagues. Tables 4.4 and 4.5 show a 'floor' effect in both the problem solving and the computation post-test with considerably more than 5% of children scoring zero.
Table 4.4  Percentage of children achieving below 25% in the post-test

<table>
<thead>
<tr>
<th>Computation Post-test Scoring Below 25%</th>
<th>Problem Solving Post-test Scoring Below 25%</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.72</td>
<td>21.88</td>
</tr>
</tbody>
</table>

Table 4.5  Percentage of children achieving zero in the post-test

<table>
<thead>
<tr>
<th>Computation Post-test Scoring Zero</th>
<th>Problem Solving Post-test Scoring Zero</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.42</td>
<td>19.14</td>
</tr>
</tbody>
</table>

It can be concluded that in this sample the proportion of children experiencing difficulty shown by a zero score was over 7% for computation and over 19% for problem solving.

4.2.2  Computation answers and problem solving answers

Hypothesis three was that children would perform worse on the post-test problem solving than on the computation questions because of the additional language factor. It was decided to scrutinize the post-test results separately for each teacher in view of the apparent different teacher style of questioning which had emerged (Part one).

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Mean Computation</th>
<th>Standard Deviation</th>
<th>Mean Problem Solving</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>teacher 1</td>
<td>44.87</td>
<td>22.98</td>
<td>48.72</td>
<td>27.05</td>
</tr>
<tr>
<td>teacher 2</td>
<td>85.00</td>
<td>20.82</td>
<td>51.79</td>
<td>41.90</td>
</tr>
<tr>
<td>teacher 3</td>
<td>45.83</td>
<td>23.70</td>
<td>52.78</td>
<td>27.66</td>
</tr>
<tr>
<td>teacher 4</td>
<td>67.71</td>
<td>27.07</td>
<td>60.42</td>
<td>36.05</td>
</tr>
<tr>
<td>teacher 5</td>
<td>75.00</td>
<td>29.57</td>
<td>25.00</td>
<td>29.57</td>
</tr>
<tr>
<td>teacher 6</td>
<td>63.54</td>
<td>36.85</td>
<td>71.88</td>
<td>35.59</td>
</tr>
<tr>
<td>teacher 7</td>
<td>77.67</td>
<td>25.18</td>
<td>57.29</td>
<td>30.82</td>
</tr>
<tr>
<td>teacher 8</td>
<td>58.33</td>
<td>33.51</td>
<td>66.67</td>
<td>35.10</td>
</tr>
<tr>
<td>teacher 9</td>
<td>66.67</td>
<td>28.69</td>
<td>40.48</td>
<td>31.89</td>
</tr>
<tr>
<td>teacher10</td>
<td>47.44</td>
<td>26.95</td>
<td>52.56</td>
<td>26.95</td>
</tr>
</tbody>
</table>

Table 4.6  Mean percentage correct by teacher
Children's computation and problem solving data was compared in a mixed plot analysis of variance. The repeated measure was computation and problem solving scores in the post-test. The children in the following three classes showed significant differences, (at a probability level of less than .005 using the Bonferroni adjustment, Pallant, 2001) between proportions correct for computation and problem solving post-test results for teachers 2, 5, and 9.

<table>
<thead>
<tr>
<th>Teacher</th>
<th>df</th>
<th>Computation Mean</th>
<th>Computation SD</th>
<th>Problem solving Mean</th>
<th>Problem solving SD</th>
<th>F Value</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher 2</td>
<td>26</td>
<td>85.00</td>
<td>20.82</td>
<td>51.79</td>
<td>41.90</td>
<td>23.81</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Teacher 5</td>
<td>26</td>
<td>75.00</td>
<td>29.57</td>
<td>25.00</td>
<td>29.57</td>
<td>56.4</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Teacher 9</td>
<td>26</td>
<td>66.67</td>
<td>28.69</td>
<td>40.48</td>
<td>31.89</td>
<td>20.29</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Table 4.7 Classes with significant differences between computation and problem solving scores.

The differences between computation and problem solving percentages correct did not reach the required significance level in the classes of teachers 1, 3, 4, 6, 7, 8, 10.
<table>
<thead>
<tr>
<th>Teacher</th>
<th>df</th>
<th>Computation Mean%</th>
<th>Computation SD</th>
<th>Problem solving Mean%</th>
<th>Problem solving SD</th>
<th>F Value</th>
<th>Significance Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>teacher 1</td>
<td>24</td>
<td>44.87</td>
<td>22.98</td>
<td>48.72</td>
<td>27.05</td>
<td>.335</td>
<td>&gt;.005</td>
</tr>
<tr>
<td>teacher 3</td>
<td>22</td>
<td>45.83</td>
<td>23.70</td>
<td>52.78</td>
<td>27.66</td>
<td>1.04</td>
<td>&gt;.005</td>
</tr>
<tr>
<td>teacher 4</td>
<td>22</td>
<td>67.71</td>
<td>27.07</td>
<td>60.42</td>
<td>36.05</td>
<td>5.03</td>
<td>&gt;.005</td>
</tr>
<tr>
<td>teacher 6</td>
<td>22</td>
<td>63.54</td>
<td>36.85</td>
<td>71.88</td>
<td>35.59</td>
<td>1.13</td>
<td>&gt;.005</td>
</tr>
<tr>
<td>Teacher 7</td>
<td>22</td>
<td>77.67</td>
<td>25.18</td>
<td>57.29</td>
<td>30.82</td>
<td>4.76</td>
<td>&gt;.005</td>
</tr>
<tr>
<td>teacher 8</td>
<td>22</td>
<td>58.33</td>
<td>33.51</td>
<td>66.67</td>
<td>35.10</td>
<td>1.114</td>
<td>&gt;.005</td>
</tr>
<tr>
<td>teacher 10</td>
<td>24</td>
<td>47.44</td>
<td>26.95</td>
<td>52.56</td>
<td>26.95</td>
<td>.480</td>
<td>&gt;.005</td>
</tr>
</tbody>
</table>

Table 4.8 Teachers whose post-test scores did not show a significant difference

The following histograms, figures 4.5 to 4.7, show the distribution of scores in those classes where the difference reached significance. The results of the ten classes may be seen in Appendix 3.
4.3 **Relationship of questioning style to post-test scores**

The first of the research questions was the relationship, if any, between teaching style and children's post-test scores. Non parametric analysis was used, as the sample size was small, and the size of groups was different. The independent variable was teacher group A or B, and the dependent variable was post-test scores. The Mann Whitney test, one of the more powerful of the non-parametric tests (Siegel, 1956) was used. Comparison of pupils' post-test scores revealed no differences in children's scores on either computation or problem post-test scores between Teaching Groups A and B:

\[
Z \text{ for children's scores on computation, was } -1.02, \quad p = 0.305, \quad p<0.05
\]

\[
Z \text{ for children's scores on problem solving was } -0.34, \quad p = 0.732, \quad p<0.05
\]

The sample size was small, N=10 and after consideration of Howell's discussion of sample size and correlation (Howell, 1982) it was decided that such a sample was too small to allow general conclusions to be drawn from a correlational analysis. However, it is worth noting that scrutiny of the scatterplots indicated a possible relationship between class management questions and computation post-test (see below figure 4.8)
While this is of interest it is necessary to use a much larger sample to satisfactorily investigate the relationship between the distribution of teacher question types and children's learning.

4.4 Scrutiny of errors

The errors in the post-test were scrutinised to see if light could be thrown on what strategies the children had adopted. As Boaler does (Boaler, 1997) the errors of the children are shown to demonstrate the lack of understanding of how to select and apply strategies. Ashcraft argued that the scrutiny of errors, particularly regularity or lawfulness in errors give an indication of the strategies used, and for the same reasons mathematics recovery schemes scrutinize the errors of children (Steffe, 1992a; Dowker, 2001).

It was decided to select two parallel classes in the same school where the teachers had covered the same content; this allowed scrutiny of errors in two different classes.
where the school context and the lesson content was constant. The classes were not set for ability and were the same age, 8 to 9 years old. The two teachers had planned their lesson together but their approaches varied, the second teacher using larger numbers. Both were teaching multiplication of two and three digit numbers using decomposition of large numbers and doubling. Both reminded their classes of the relationships within the numbers and taught multiplication of 2 digit numbers using two methods, doubling, and decomposition. Both taught the relationship of multiplication to addition and worked through examples of single digit numbers, such as

'What is another way to write 8+8+8+8+8'

Children transformed sums like this on the white board and gave the answer. Text boxes 4.1 and 4.2 show similar post-test questions for the two classes and errors.

<table>
<thead>
<tr>
<th>Text Box 4.1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>How else can you write 3+3+3+3+3 and what would be the answer?</strong></td>
</tr>
<tr>
<td>1. 3 x 5=8  girl</td>
</tr>
<tr>
<td>2. 4 x 3=16  girl</td>
</tr>
<tr>
<td>3. 33+33=99  boy</td>
</tr>
<tr>
<td>4. 3 x 3=6  boy</td>
</tr>
</tbody>
</table>
Bearing in mind that the teacher had taught the relationship between addition and multiplication, in the first answer, the girl appeared to have understood that it was important to note the number of threes, and has written the sum correctly, but in the actual computation has used addition instead of multiplication. Answer two is close to the correct one and it may be speculated that this is a computation error. Answer three indicates a lack of understanding of place value, and of addition, and the answer four has used a multiplication sign but carried out an addition sum. If we judge by the results of the post-test, in this class approximately 22% of the children had not understood the relationship of multiplication to addition using single digit numbers.

The questions tackled in the class where the teacher focused on larger numbers were 8 times 14, and 879 times 23. In one case the class worked through the problem by doubling, and in the other by decomposing 23 into 10, 10 and 3, the teacher drew a grid on the white board:

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10 x 879</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 x 879</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 879</td>
<td></td>
<td></td>
</tr>
<tr>
<td>total</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Some of the questions set in the post-test were
Write down how you would answer 4 times 13

1. What is the answer?
2. Write down how you would answer 8 times 17?
3. What is the answer

The children demonstrated considerable confusion over which strategy to use in the post-test. For example, H, a year 4 boy, wrote the following when asked to multiply 4 by 13 and explain how he had done it:

TEXT BOX 4.3 Errors in the post-test

4 times 13
make 5 into 10 and × 13 which is 130, then halve it

8 times 17
make 17 into 20 then times it

C, a year 4 girl, wrote:

4 times 13
4 × 13 = 26 × 13 = 52 × 13 = 104 add 13 = 116
8 × 17 = 94 8 × 10 = 80 × 17 = 14 that = 104
Both children had realised that numbers are changed but not how or why. H realised a useful strategy was to use 10, or 20 but not how to do this. C, the girl, had started by doubling, reached the answer, but continued to double and add another 12 (possibly she meant to add 13).

Other errors in the post-test are shown below:

(4 x 13)
1. 4 x 13, x 2 x 4 x 8 x 16, take 2 away (girl)
2. 4 x 10 = 40, add 3 which is 43 (boy)
3. 4x13=43 (girl)
4. 4x13=12 (girl)
5. 4x13=66 (boy)
6. 4x13=86. 10, then 13 4 times. (boy)
7. Make 5 into 10 then times 13 which is 130, then halve it (boy)

For 8 x 17
1. I added 4 x 10, which is 40, and 4 x 17=79 (boy)
2. (8 x 17) used a grid for 8 x 17 (girl)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>8 x 10</td>
<td>80</td>
</tr>
<tr>
<td>80 x 2</td>
<td>160(answer)</td>
</tr>
</tbody>
</table>

The first girl has understood that it is useful to decompose and to double. Child number 2 decomposes 13 into 10 and 3 but does not understand that both numbers have to be used to multiply 4 and then added. Answer 4 appears to indicate no understanding of the second element of numerosity, the understanding of the relationship between numbers (Butterworth, 1999; Nunes, 1996). Answer 6 decomposes 13 into 10 and then into 13 but does combine multiplication and addition. Answer 7 appears to give a strategy for arriving at an approximate figure but does not include acknowledgement of the 2 x 13 in this overestimate; this forgets
to subtract the 8 (2 x 4) from 10 (2 x 5). For 8 x 17 the second answer decomposes, carries out the first part of the sum, 8 x 10 but then doubles the answer, instead of adding 7 x 8 to the answer.

It cannot be claimed that there is a regularity or lawfulness in these errors (Ashcraft, 1992). If many of the children had made exactly the same error it might have indicated the use of a certain strategem or model of calculation. However, it is argued that the children's errors show confusion over the applications of strategies, and what Boaler (1997) referred to as rule following behaviour. The pupils have learned a rule and applied it wrongly. This will be considered further in the Discussion, Chapter 5.

The errors indicate that the children have not understood the relationship between the numbers, the second element of numerosity (Nunes, 1996; Butterworth, 1999). In particular they have not understood the relationship of addition to multiplication, not had they understood how the methods demonstrated by the teacher related to the problems they were asked to answer. A significant number were not able to answer the conceptual questions, requiring analysis of the problem and correct choice of procedure.

4.5 Research questions and conclusions

4.5.1 Part one: teachers' questioning

Hypothesis one could be accepted. Teachers asked more computation questions than problem solving questions. Scrutiny of the data for any systematic pattern of variation in question type, i.e. teaching style, showed variability, but two questioning types emerged although caution should be exercised in generalising from a small
sample. The first, called type A, consisted of seven teachers who asked a greater proportion of computation and problem solving questions than factual and class management; in other words, they asked a greater proportion of mathematics questions than the second group, type B, three teachers, who asked a greater proportion of factual and class management questions. The difference was particularly noticeable in the proportions of factual and problem solving questions. Type B asked twice as many factual and half as many problem solving questions.

The research question concerning gender indicated no significant differences in the number or type of questions directed at boys and girls, nor was there a difference in post-test performance.

The sample was too small to indicate a significant relationship between teachers' use of different question types and children's post-test performance although it appeared that the number of class management questions may be a predictor for computation performance in the post-test.

4.5.2 Part two: children's learning: post-test performance

The second hypothesis was that, in keeping with estimates (Butterworth, 2002, Snorre 2002) it was expected that between 2% and 10% would not have been able to access the learning in the whole class session and that this would be reflected in the post-test. Scrutiny of the scores in the post-test showed that over 7% scored zero on computation post-test answers, and over 19% scored zero on the problem solving items. These were taken from material taught in the preceding half hour, indicating that in each class there were children who had not understood most of the content of the preceding lesson, and that nearly 20% of the children had significant difficulties with problem solving questions.
Hypothesis three, that children would perform better on computation questions than on problem solving questions could be accepted for only three of the ten classes.

Scrutiny of the errors indicated a lack of understanding of the relationships between numbers, and how the method or strategy reflected this relationship, the second element of numerosity (Butterworth, 1999; Nunes and Bryant, 1996). The error analysis was particularly relevant to the research aim of considering the difficulties children experienced in mathematical questions.
Chapter 5

Discussion

5.0 Discussion

The first part of the discussion will relate to the research aim of exploring types of question, and style of interaction in the classroom: the second part will consider the learning of the children in this study. The division between teaching and learning is to some extent artificial; for example, the discussion of the relationship between class management questions and post-test computation is in both parts one and two. It will be included in the first part but has obvious implications for the research question about pupil learning, discussed in part two.

The third section of this chapter returns to the research question of the validity of higher and lower order questioning raised in the Literature Review, and constructivism; the final part considers issues for the theory and practice of Educational Psychology.

5.1 Part 1 - Teacher questioning

5.1.1 Computation and problem solving question types

Teachers asked a greater proportion of computation than problem solving questions, although the percentages were close, and well above factual and class management categories. The majority of questions in this study were not housekeeping or class management. Table 5.1 shows a greater proportion of computation questions in this study than closed questions in the ORACLE study (Galton et al., 1999) and a greater proportion of problem solving questions in this study than open questions in the ORACLE study.
Table 5.1 Teachers’ question types compared to the ORACLE findings

Table 5.1, above shows that there is a smaller proportion of class management questions in this study than task supervision in the ORACLE study; if the routine and task management questions in the ORACLE study are collapsed to form one group, reflecting the class management question category in this study, the 47% in the ORACLE study is considerably more than the 12% in this sample. It is possible to conclude from this that these ten teachers spent considerably less time on task supervision and routine management during this part of the lesson, than did the teachers of the ORACLE study. It should be remembered, however, that the ORACLE study was based on time sampling throughout the day, however, while this study focused on one short period of planned whole class teaching.

There are some difficulties in comparing the Numeracy questions in this study with the ORACLE question types because of several factors. Firstly, the ORACLE question types, open and closed, applied to all subjects, not just mathematics.
Secondly, it is argued here that the closed and open categories of the ORACLE study are superordinate categories in which other types of questions fit; it is argued that recall, fact and computation question types are all closed question types but are further defined by the kind of thinking they may evoke. Additionally it is argued that the problem solving type of questions in this study may be either open or closed (see below) and so does not map on to the open question type of the ORACLE study, defined by the teacher seeking more than one answer. The ORACLE study used a more stringent definition of an open question than other researchers have used: the teacher had to demonstrate that the question was open by looking for more than one answer:

Teacher: ‘What do you think is happening when the solution turns blue?’

Pupil: ‘Copper’

Teacher: ‘What else could it be? What do others think?’

In this study of questions in the numeracy session, teachers did turn to the rest of the class but were more often likely to ask: ‘Is he right? What do you think?’ In other words the teacher was looking for confirmation or disconfirmation of an answer, rather than for several alternatives.

A more useful categorisation for mathematical questions appeared to be the recall, computational, and problem solving question types used here. Computational and problem solving question types mapped more easily on to Boaler's definition of procedural, and conceptual question types used for secondary school mathematics (Boaler, 1997) than on to the ORACLE question types. However, it should be noted that Boaler's definition of procedural subsumed within it both computational and fact/recall questions (see Table 1.3 in Chapter 1, and Table 5.2, below). It is argued that the subdivision of Boaler's procedural type into factual recall and computation is
relevant here because of the discussion of the role of recall of familiar number facts (Ashcraft, 1992) raised in Chapter 1.

<table>
<thead>
<tr>
<th>Author</th>
<th>Type (closed)</th>
<th>Type (open and closed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boaler (1997)</td>
<td>Procedural</td>
<td>Conceptual</td>
</tr>
<tr>
<td>Galton et al., Oracle (1999)</td>
<td>Closed, routine, task supervision</td>
<td>Open</td>
</tr>
<tr>
<td>Present study</td>
<td>Fact/recall</td>
<td>Computation Problem solving</td>
</tr>
</tbody>
</table>

Table 5.2 Differences in question classification

Both Butterworth and Ashcraft have drawn attention to the role of practice and familiarity in developing mathematical skills (Ashcraft, 1992; Butterworth, 1999), and the Numeracy Strategy (1999) has emphasised the role of practice and familiarity with number facts so that for this study, it was judged important to identify which questions were aimed at practice and increasing this familiarity with number facts.

5.1.2 Questioning style

Although a research aim was to explore types of teacher question in the teaching of mathematics, variability in teacher questioning had been predicted from the pilot study, and the data were scrutinised for any systematic variation, the clustering found in this study was unexpected. The difference in the two groups appears to lie in the greater proportion of both computation and problem solving questions in group A, the group of seven teachers, and the greater use of factual and class management questions in group B.
The difference was particularly noticeable in factual questions where group A teachers asked well under half the percentage of group B teachers, and in problem solving questions where group A asked twice the percentage of group B. It is not clear whether these findings may be generalised to a larger group, nor what the impact of the different styles may be on learning. The two groups may map on to research identifying two groups such as that of Newstead (1998) discussed in Chapter 2 and of Saxe et al., (1999); these researchers divided their sample of teachers into a group aligned with traditional practice and a group aligned with reform practices; the reform group spent more time on pupil led problem solving (Saxe et al., 1999). However, the selection of the members of these groups was, they admitted, done subjectively, not on questioning practice. In the study by Newstead (1998) it was also observed by the author herself that the classification of teachers into two different groups was at least partly subjective.

It might be expected that those teachers who spend a greater proportion of their questioning time on mathematics questions would have a greater effect on mathematics learning. However, contrary to what might be expected, the number of class management questions related to better performance in the computation items in the post-test. One speculation from this finding is that those teachers who asked a greater proportion of class management questions were more aware of those children who were not engaged, nor on task. It may also be argued that a greater proportion of class management questions may have reflected a more difficult class. To be able to conclude that the proportion of class management reflected a teacher characteristic, rather than a class variable, it would be necessary to sample different teachers' questioning with the same class, under as similar conditions as possible.

There appeared to be no particular characteristic of the three teachers in the smaller group who asked a smaller proportion of mathematics questions, although two points
are worth noting. The first is that two of the three spent considerably less time than is usual on the teaching part of the Numeracy session – 14 minutes and 16 minutes. This is noticeably less than others in this sample. Only one of the larger group A spent less than 20 minutes, and the other 6 spent between 23 and 33 minutes. The mean time was 24 minutes and if the Numeracy Hour is divided into three roughly equal parts it might be expected that the first part should usually be about 20 minutes. It may usually be the case that it is over one third of the lesson as in classes observed for the whole numeracy session, the plenary part was considerably shorter than the other two sections, averaging around 10 minutes. It is not possible to speculate on reasons for those teachers spending so much less time on the introduction to the mathematics topic, but a valuable source of data here would be the teachers’ own reflections on whether this was usual practice for them, and possible reasons for their allocation of a shorter period to this part of the lesson than is usual.

The two groups in this study were not predicted, but it is argued that it may be helpful to find out if it is possible to generalise from this result by looking at a much larger sample to see if these distinct questioning types are replicated, and controlling for the level of special educational needs. It would be of interest to see if the level of special educational need, and time spent on the whole class teaching, were in any way related to teaching style and to learning. Teachers and educational psychologists might pursue the two types as a spectrum or continuum for considering different styles of whole class teaching, and for use with video feedback for teachers engaged in self evaluation.
5.1.3 Unknown variables intervening between question and answer

It is argued that there are at least three intervening variables between question and answer which relate to any analysis of question type and learning. The first is the teacher's intention in asking that question, a functional definition of question type, 'What do I want this question to do?' (Morgan and Saxton, 1994, p.41). The second variable is the child's understanding of the question and that will be discussed below in considering the part played by language. The third variable, perhaps harder to investigate, is the nature of the thinking done by the child to answer the question and this will be discussed further in Chapter 5, Section 2.1 on children's mathematical thinking and learning.

Teacher intention was not always easy to determine from the question. An apposite example of the difficulty in determining teachers' intentions and the child's thinking is given below. This question was one which a teacher had taken from the Numeracy Strategy where it was given as an example of an open question:

<table>
<thead>
<tr>
<th>TEXT BOX 5.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>pupil</td>
</tr>
<tr>
<td>Teacher,</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

The first answer offered was, it might be argued, an answer to a problem solving question – how could these numbers be combined? A number of strategies could be offered and answers compared to make the biggest number. The boy who answered
had treated this as a problem to be solved and had thought about place value in a problem solving way, to answer a question which the teacher then rephrased as a computation question, or a question requiring a series of computations, if the pupils had time to carry out the operations and compare the answers. If the pupils considered what was likely to emerge from using various operations, then it might be considered a problem solving question. If we take Boaler’s definition of questions being either procedural or conceptual, (Boaler 1997) it could be argued that the child answered this question in a conceptual way, but that it was intended to be procedural. Further examples of what appeared to be problem solving questions were:

TEXT BOX 5.2

1. ‘What relationship has 10 to 20?’
2. ‘How can we multiply by 20?’
3. ‘Can anyone tell me some easy ways of multiplying by 5?’

In the first question the teacher wanted only one answer, that 10 was half of 20, as his reply was, ‘No’ when a child said: ‘It’s smaller than 20.’ The question was, in fact, a computation (closed) question, or if the child had the fact that 10 was half of 20 in memory, it may have been a factual/ recall answer. Had the teacher taken other responses, such as the first offered, then it would have been an open (ORACLE definition, Galton et al., 1999) or problem solving question.

The second question, ‘How can we multiply by 20?’ appeared to be a problem solving question but was answered as recall of a rule: ‘You put a zero on and double it.’ This was the only answer taken, and so that was in reality, a recall/fact answer.
The third question, ‘Can anyone tell me some easy ways of multiplying by 5?’ was also answered as a rule recall: ‘Times by 10 then halve.’

All these questions which appeared to be problem solving, or conceptual questions, were in fact, closed questions, requiring either recall of a rule or number fact, or computation. Furthermore, Boaler has pointed out that what started as an open question may be redefined by the teacher as a series of narrower, closed questions. If the student could not answer a question the teacher would tend to pose it as a series of steps of multiple choice questions: ‘Well, is it 4 or 5? Is it length or width?’ (Boaler, 1997, p. 22)

From the questions above it appeared that more of the questions were closed questions to which the teacher expected one answer, the answer in the teacher’s mind, than the question itself appeared to be. What might seem to be an open question was often not, as the teacher would expect one answer, or move on after only one answer. This may have been due partly to the expectation of pace required by the DfEE in the Numeracy Strategy, (DfEE, 1999) and possibly partly to the teacher’s wish to keep the lesson moving, rather than wait for alternative answers. The issue of pace or lesson speed as a characteristic of mathematics lessons is further discussed below (Boaler, 1997). Additionally, there may have been an effect from the experimental situation; the teacher may have been less willing than usual to wait for alternative answers because of the presence of an observer.

For the numeracy session the ORACLE (Galton et al., 1999) definition of an open question was unsatisfactory for the purposes of considering the child’s thinking. The nature of problem solving questions in this study was closer to the conceptual category of Boaler (1997). It did not appear a necessary precondition for a problem solving question to have more than one answer; any question which required the
child to consider which mathematical procedure was the appropriate one was a problem solving question, whether or not it had more than one answer.

5.1.4 Gender and questioning

There were no differences in gender and questioning practices of teachers. However some individual differences were observable in teacher practices and in this area qualitative research may be useful. It is more likely that, using in depth analysis of one teacher’s questioning, it would be possible to show whether he or she questions boys and girls differently. For example, one teacher asked more questions of all types of boys, and another teacher asked more questions of all types of girls. The argument to support qualitative analysis of the interaction in one class would be practical in application for teachers’ own self evaluation; a teacher is likely to wish to consider his or her own teaching in depth, and how factors in the class and school affect the interaction. Furthermore, from the point of view of the child, it might be argued that a study demonstrating the overall lack of gender difference in teacher/pupil interaction, is irrelevant to the children in the classes where there were differences in questioning towards boys and girls.

In this study there did not appear to be gender differences shown in teacher questioning, and it would be difficult to establish that teacher differences observed were related to gender per se, rather than other factors; implicitly, the teacher may be aware of the preferences of some children for a lower level of social exposure, (Newstead, 1998) both in questioning and reinforcement, and may be taking into account the learning and behavioural difficulties of various children.
5.1.5 Class management questions, and other teaching variables

An unexpected indication was that class management may be related to performance in the computation items. In fact, class management questions were not part of the planned investigation. In speculating why there may have been a relationship between class management questions and performance the first possible reason is that it may be a quirk of this particular sample. Although question numbers were large the sample of teachers was small and teachers who asked proportionately more class management questions may have had a greater effect on the data than in a larger sample. However, if this were to be the case in a larger sample, it raises some interesting questions. Class management questions were usually directed at any child or pair whose attention was perceived to be not on the task. Examples are shown in the text box below.

TEXT BOX 5.3

'Would you be better sitting here R?'
'Can you see from there W?'
'Could you stop fiddling with that J?'
'Could you look at the white board please?'
'Could I have everyone's attention?'
'What are you two messing about with?'

The researcher had not originally intended to analyse these as they have not always been categorised as questions (Galton et al., 1999; Scarth and Hammersley, 1986). They do not, as a rule, demand an answer, but are more usually rhetorical ways of maintaining the attention of the entire group. However, these questions may be of interest; if a greater number of class management questions were to be related to
better performance on the post-test, and it was assumed that this reflected the learning which has taken place during that numeracy session, then perhaps those teachers who made a greater number of comments directed at class management, increased the likelihood of learning. It was interesting to see that the teacher who asked the highest number of class management questions and who also spent the shortest time teaching had high post-test results – a mean of 60% correct. It might be speculated that this teacher was aware of who was not on task and more often than other teachers used questioning to bring their attention to the task. There are alternative explanations which should be considered. Less mathematical material may have been presented which in turn may have led to more efficient learning of what was presented. It would be of relevance to children’s learning to establish the amount of material which may be presented in any one session, and still result in the effective learning of most children.

The issue of how much material may be introduced to children in one session relates to time spent in whole class teaching. Although the numeracy session was originally intended to be prescriptive in both its content and structure it was noticeable that there was considerable time variation in these lessons. The variability in length of this section of the numeracy lesson was mentioned above, from 33 minutes to 14 with the mean time being 24 minutes. The relationship between time and learning was not straightforward, illustrated by the teacher referred to in the previous paragraph; this teacher had the shortest whole class teaching time and the greatest number of class management questions, and the class achieved a mean of 60% of correct answers in the post-test, at the top end of the range.

It is also necessary to note that time spent may not have a direct relationship with the amount of material presented. Pace of teaching is an intervening variable (it is possible that less time reflects a faster pace of teaching) and pace is yet another
variable which has been noted as a feature of questioning in mathematics. Boaler (1997) found in secondary schools that some students complained about the pace being too slow and others complained (predominantly girls) that it was too fast:

‘For the teacher it shows how difficult it is teaching a group at the same pace even when they are meant to be of homogeneous ability’ (Boaler, 1997, p.33)

Boaler (1997) observed teachers moving quickly round the class until they got the right answer and noted that wrong answers were not a trigger for analysis. Boaler has concluded from this, that students gain the idea that mathematics questions should be answered quickly and that speed of thinking rather than depth of thinking is a feature of mathematics. It is worth contrasting this with the emphasis placed by Butterworth and Ashcraft on repeated practice so that pupils become secure and familiar with number facts and build up strong associative networks (Ashcraft, 1993; Butterworth, 1999). Comments from the students interviewed by Boaler reflected this:

‘The teacher rushes through methods faster than most pupils can cope’
(Boaler, 1996, p. 30)

It may be speculated that the number of class management questions reflected the variable identified as ‘with-itness’ or the ability to spot potential trouble spots. Kounin (1970) identified this in his analysis of 80 primary age class rooms; successful teachers who showed this ‘with-iness’ had fewer instances of misbehaviour. In the Numeracy Strategy the teacher is likely to be restricted in movement as the strategy places considerable emphasis on the use of visual learning strategies such as the number line or number grid, and the working out of calculations on the white board. It might be speculated whether there is additionally
an 'action' zone (McKee and Witt, 1991), in which the majority of the interaction takes place. This is traditionally the students at the front and centre:

'who are recipients of more teacher directed verbal interaction, questioning and eye contact' (McKee and Witt, 1991, p. 827). Where the teacher is tied to one position in the class the authors made the point that, 'the two way flow of instruction is most unequally distributed' (ibid., p. 827)

McKee and Witt (1991) have made the point that different seating arrangement may suit different presentations and strategies; this resembled the argument put forward by Galton et al., (1999) that the true art of pedagogy lay in matching the strategies to what was to be taught to ensure the most effective learning. However, they have said: 'the traditional straight row arrangement appeared to be most effective for traditional lesson presentation' (McKee and Witt, 1991, p.828)

Furthermore they found that density of seating has most effect on performance where there was a high information processing load (McKee and Witt, 1991); it is argued here that various sequential stages of calculation in mathematics require this, and so seating as a part of class management is a factor to be considered in the Numeracy lesson.

Of the ten classrooms in this study, in eight the children sat on the floor in front of the white board. In two they sat at their usual tables which were arranged so that every child could see the white board. It may be that such density of seating in those 8 classes over a relatively long period – up to 33 minutes – contributed to the difficulty of keeping children on task. For example, one child, J, who scored zero in the post-test, had a Statement of Special Educational Needs, sat at the back of the group by the book case and the video recording shows him taking books out throughout the
session to look at pictures. The number of class management questions was low in this class, and the child did not impede the learning of others, but such an observation raises questions about how to implement inclusion and ensure the learning of all children in the numeracy lesson.

5.2 Part 2 - Children's Learning

5.2.1 Children's processing of questions

While to some extent the teachers' intentions can be surmised through what followed from the question, whether they actively looked for a further answer, or prompted the pupil to elaborate, it is more difficult to establish the second variable referred to above, - the nature of the child's thinking evoked by the questions. This issue was raised in Chapter 1, Section 4, and in Chapter 2, Sections 3 and 4. Askew and Wiliam (1995) noted this difficulty when they gave an example of a number bond such as 5+5 being known by heart (tie questions are more usually familiar number facts to be retrieved) and 5+6 being 'figured out', or calculated. The difficulty in knowing exactly how a child has reached an answer became apparent in the pilot study and is discussed above; the researcher had classified questions such as '3 add 4' as computation, while one of the other raters classified them as recall. The researcher assumed that some children of eight would use counting strategies for this question and the rater assumed that such number bonds would be a fact to be retrieved from memory.

It is acknowledged that while this is an issue underlying analysis of question types, and difficult to resolve, some clues may be used to point to a child's thinking. One way to find out how the child did the question is to ask what strategies s/he used as has been done by Siegler, whose work is discussed below (Siegler 1987a, 1988b).
Questions in the post-test in this study asked children to describe the steps they took but it is not always possible for them to remember or write down these steps. Siegler's practice of asking the child to describe what s/he did appeared most useful for accessing thinking but practically this would have been difficult to do with more than a few children in this case. It may also be argued that such methodology is of questionable ecological validity, in the context of scrutinising the learning in the numeracy lesson. However, the mathematics recovery systems (Dowker, 1999, 2002; Wright, 1992) advocated in depth analysis of each child's thinking, in order to draw up a programme of learning.

It is argued that speculating about the way in which children process the questions is central to the research question of understanding difficulties with learning. It may be that children's learning takes different forms; Siegler (1987a, 1988b) studied the arithmetic strategies in addition of 7 year old children and found that 44% used basic retrieval from memory, 38% used counting, 9% used a decomposition strategy, and only 1% used the very inefficient strategy of counting on by ones. In Siegler's model there was a trend to greater reliance on retrieval as the child grew older so that by 8 or 9 he argued that the bulk of processing for the basic facts of addition and multiplication was generally retrieval as these had become well known and familiar facts held in memory rather than calculations to be worked out. The children in this study, who were in the majority, 8 and 9 years of age, may have been moving from a counting on or calculation strategy to reliance on memory for familiar computational questions. It seems possible that some children would be relying on retrieval and some on calculation and this may account for the finding that with unlimited time most children can add accurately (Jordan, 2002). This area appears worthy of further research particularly in relation to those children with mathematical difficulties, in considering a delay or difference model (Snorre, 2002, discussed in Chapter 2). Snorre argued that for some children with mathematical difficulties it is a question of
delay, and these children catch up, while those with persistent mathematical difficulties he called mathematically disabled. It might be speculated that certain children have difficulty in embedding the facts in memory, and so take longer to move to a retrieval mechanism. They may be calculating the answer and need concrete aids to help this. The time to reach the answer is likely to be longer than the time for retrieval of a fact from memory and may lead to a higher error rate.

To attempt to access the mathematical processing of the children in this study, errors in the post-test were scrutinised. The study of errors may give clues to how children tackle mathematical questions (Ashcraft, 1992). For example, if there were a small calculation error this might point to computation, rather than recall of a well memorised number act. If the answer to a question of multiplication is another multiplicand of that table, for example, $3 \times 4 = 16$, then this may indicate the use of memory of a table (Ashcraft, 1992). This will be discussed further in the section below considering children’s difficulties with mathematical thinking.

5.2.2 Children’s processing: computation questions

Before discussing the results of this study it should be noted that the demands of the situation – the observation of classes in a situation resembling usual practice, as far as possible, - made it difficult to control for levels of difficulty of material presented. It was noticeable that in the same school two parallel classes were tackling problems of differing difficulty. Both were covering multiplication but while one was covering it with small figures and finished with slightly larger ones, the other teacher went rapidly on to much larger figures. Bearing in mind the greater difficulty experienced with calculations concerning larger numbers (Ashcraft, 1992) and the theories discussed in the literature review relating to possible different ways of dealing with large numbers (Molke, 2002) it was not unexpected that the class given larger number
questions showed poorer performance in the post-test. Computation items were done better than the problem solving questions, as was predicted. However, nearly 12% of children scored below 25% and 7% of the whole sample scored zero.

<table>
<thead>
<tr>
<th>Computation Post-test Below 25%</th>
<th>Problem Solving Post-test Below 25%</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.72</td>
<td>21.88</td>
</tr>
<tr>
<td>Computation Post-test Scoring Zero</td>
<td>Problem Solving Post-test Scoring Zero</td>
</tr>
<tr>
<td>7.42</td>
<td>19.14</td>
</tr>
</tbody>
</table>

Table 5.3 Percentages of children scoring below 25% and scoring zero

Using these percentages as criteria indicated levels of difficulty above the estimates of both Butterworth (2002) and Snorre (2002) for children with difficulties in mathematics. There may be variables in this study which partly account for this. Firstly, the estimates of Butterworth and Snorre were based on a more heterogeneous catchment area than was the case in this study; the catchment areas of these schools have been defined by the Local Education Authority as areas of high need and poverty using such indicators as numbers of single parent families, numbers of children entitled to free school meals, numbers of unemployed adults. All except one served large local authority housing estates.

Secondly, the criteria used here by the researcher of twenty five per cent correct may have been too high. Other studies have used the children scoring below the fifth centile (Shalev et al., 1998). Because of this the researcher adopted the more stringent criteria of scores of zero; this may more accurately reflect a conservative estimate of the number of children with significant mathematical difficulty.
percentage of children scoring zero in computation was 7.4. A total of 19 children in the sample of 256 scored zero on questions which were directed at material just learned, and which were very similar to questions put by the teacher in the preceding 20 minutes. The question arising from this for educational psychologists, in particular, is the learning of these children in the whole class situation.

If a social constructivist position is adopted in considering teaching and learning, then questioning for each child should be within the zone of proximal development for that child, the level of understanding just beyond the current one, which the child can reach with adult or peer help (Vygotsky, 1978). Scores of zero indicated that questioning had not been accessible for these children and that even the easiest questions in the post-test were not within the zone of proximal development for these children. It is worth considering this in the light of the remark below made by the class teacher, about the child J.:

'I know J hadn't a clue what we were talking about but at least he didn't interrupt. It was far too difficult for him. He's still on number bonds to 10 so what's the point of expecting him to times double figures'

(Chapter 6, Section 1.6)

Teachers appeared to accept that the content of the lesson would be beyond the next stage of learning for some children, and the concept of mathematics recovery programmes for such children is discussed briefly below.

One of the characteristics of the classes which appeared in observation was the lack of familiarity of the children with high frequency facts. The teacher and some of the children led in rehearsal of number bonds and tables, but on various occasions when the teacher dropped out of the rehearsal, it became clear the children did not have this familiarity. One possible explanation for the high level of children scoring zeros
is that automaticity may not yet have developed in many of the children in this sample. This would affect results in the post-test, where children were given 2 minutes for each question, and the questions repeated at the end of the test with a further minute for each. However, for those children who have not yet achieved a degree of automaticity this may not have been enough time. Jordan (2002) has claimed that the vast majority of children can complete arithmetical calculation if they are allowed unlimited time.

In discussing the part played by rapid recall of well known number facts it is useful to consider in some detail Ashcraft's findings, discussed in Chapter 1 (Ashcraft 1992). He argued that rapid processing of low number arithmetic problems was a function of memory and that, analogous to the faster retrieval of high frequency words, a parallel argument in arithmetic may be expected. In general children have longer reaction times in responding to problems with larger numbers, suggesting that counting procedures take place in calculation.

Ashcraft's arguments support a growing degree of automaticity in the answering of low number calculation questions, increasing with age. It might be argued that this is analogous to reading as children move from sounding out words to reading 'by sight' automatically and quickly. The children in these classes were eight and nine years old with a percentage of children in year 3 who were seven and eight years old. In the mental mathematics sessions in which teachers rehearsed tables and number bonds it appeared that a sizeable proportion were not familiar with these number facts. It may be speculated that the teacher and those pupils who had automaticity in recall of familiar number facts led the other pupils in this session, particularly when pupils were reciting facts or tables together. It appears likely that children making errors in simple computation questions (3 x 4) have not developed sufficient familiarity with number facts to be able to use retrieval.
5.2.3 Children's processing: rule following behaviour

While there was no pre-test given in this study, the level of misunderstanding was higher than had been anticipated and warrants further discussion. Boaler (1997) in discussing the misapplication of rules argued that students believed they had to remember a rule to answer a question. A consequence of this was that they became confused when situations changed, just as the above answers showed the much younger children did. They applied the wrong rules or a mixture of two procedures. Boaler argued that pupils' behaviour was cue based, since the ways questions were tackled depended on which cues were picked up. These cues might be irrelevant cues rather than mathematical understanding of the question and mistakes in the use of rules arose because students did not know when situations were mathematically similar. Boaler observed that in a top set in secondary school 45% of the set attained lower scores in a post-test than they had attained in a pre-test; the students were motivated, 'watched, listened then practised', but for almost half of the top set, (10 of 22), 'learning was extremely ineffective and for almost half of the group it may even have been detrimental' (Boaler, 1997, p.132,133).

However, retrieval of number facts from memory does not necessarily indicate understanding of the relationships between numbers, nor that the student knows how to tackle problems. The Cockcroft report (DES, 1982) drew attention to the lack of understanding often accompanying knowledge of formal mathematics rules. An example of this is yielded by the post-test items; error analysis gave clues to the cognitive processes and strategies used by the pupils, and it was apparent that the basis for rules had not been understood. This led to the misapplication of rules, or what Boaler termed 'rule following' (Boaler, 1997, p.36). An example in this study arose from a class in which the class teacher had taught the use of alternative
strategies to multiply large numbers; she demonstrated both doubling, and decomposing large numbers into smaller numbers, to multiply two digit numbers. She worked through examples showing that 20 times a number was 10 times the number doubled. She then taught the class how to break down numbers into tens and units to multiply, and worked through examples on the white board, questioning frequently. One example was 7 times 17, which she demonstrated as 7x10, added to 7 x 7, yet, in the post-test, none of the class could do 8 times 17. Only twenty per cent (five of the class of twenty five) reached the answer for 8 times 13. When asked if 10 times 20 was the same as 5 times 40 most answered that it was not, although this question was a repetition of one asked by the teacher and demonstrated on the white board.

If the definition of numerosity is an understanding of the relationships between numbers (Nunes, 1998) then they did not show this at a level necessary for understanding how decomposition allowed the use of smaller figures in multiplication of two digit or three digit numbers. Well over half (64%) of the children in this class did not have this understanding. Using single digit numbers, 22% of a class of the same age showed they did not understand that multiplication is a way of adding the same number several times, fifteen minutes after having been taught this.

An example of both a lack of understanding of relationships between numbers (Nunes, 1998) and confusion in applying rules (Boaler, 1997) follows. The class teacher had spent ten minutes on how to add numbers which were 99, 199, 299, using a number grid and rounding up the numbers to hundreds. One answer to this post-test question was, $199 + 199 = 2819$. It may be speculated that he had added from left to right following certain addition rules. What is apparent is that he had not grasped the relationship of 199 to 200.
It is difficult to explain why so few children were able to tackle questions in a post-test, which they had worked through as a whole class. One explanation may be pace of teaching. Boaler noted that teachers worked too quickly for most pupils (Boaler, 1997), or it may be linked to the failure to allow children to construct their own learning. Boaler argued forcefully that the nature of the pedagogy of mathematics teaching did not as a rule allow students to construct their own learning. Where students had been allowed more 'open ended learning' exam results were better. For example retention of learning rose to 36% for statistics the students had used in their own projects (Boaler, 1997).

A question arising from this confusion over the application of different methods is whether giving all children alternative ways to work out arithmetical sums is always useful. It is possible that for some children the teaching of alternative strategies allows them to make connections, but in the case of children who are less familiar with the relationships between numbers, they may retain the rules but be unable to apply them with understanding. For example they remember that the numbers are made smaller in decomposition, and they remember they must add something but they are not sure what. If it leads to confusion for some children, but enriches the understanding of number for others the question of benefit for all children arises. For those children scoring zero in either computation or problem solving items, it may be worth considering the possibility of a recovery type of programme. Steffe and his colleagues have used one in Australia, parts of the United States, and parts of the United Kingdom (see Steffe, 1992). Another recovery programme has been trialled in Oxford (Dowker, 2001). For children for whom the questioning is beyond their zone of proximal development (Vygotsky, 1978) such recovery programmes, if properly evaluated, may serve to establish that second element of numerosity, the relationships between numbers.
It may be that different styles of teaching work for different levels of previous exposure and understanding: children were found to have an advantage in problem solving questions if they were used to problem solving in class and had a rudimentary knowledge of what was introduced – fractions in this case (Saxe et al, 1999); the same benefit was not found for children without this rudimentary knowledge.

5.2.4 The relationship between language and numeracy

It was the case that in three classes children performed significantly more poorly on problem solving than on computation questions. It is speculated that this may be because of the additional language processing required in understanding problem solving questions. For some children the additional processing of language may have affected their ability to answer. Snorre’s research, discussed in Chapter 2, (Snorre, 2002) argued that there were children who had mathematical and language difficulties, and children who had mathematical difficulties alone. The second group appeared to have more long lasting difficulties with mathematics. However, the relationship between numbers and language may not be straightforward. Campbell and Fugelsang (2001) presented arithmetic calculations to students either in digit form or written word form:

'3+4 = 7 or three plus four equals seven'

for a true or false answer. Students reported that they were more likely to use calculation with the words and retrieval from memory for the digits. Reaction times were significantly longer with the words and were also longer when the answer was correct. There was also an interaction between the size of the numbers and the presentation form so that reaction time for words increased, relative to digits, for the larger numbers. In other words, the larger the numbers, the greater the effect of word form became. Campbell and Fugelsang argued that this indicated that surface
format influenced central processing, forcing calculation rather than retrieval of a number fact. The greater effect for word format problems was also larger for correct answers and the authors argued that is because the subjects were forced to complete the calculation for the correct answers, whereas in many of the false answers it was possible to reject the answer before they completed the calculation (Campbell and Fugelsang 2001).

The variable of language appears more likely to affect problem solving questions than computation questions as problem solving questions tend to have more words and fewer numbers. This may add to the likelihood of children calculating the answer rather than retrieving a well known number fact (Campbell and Fuselgang, 2001) and raises the validity issues underlying classification. The post-test was presented verbally by the researcher and so it is possible that language had an effect on both types of question, computation and problem solving. Given the effect found by Campbell and Fugelsang the children may have been more likely to use computation or calculation for even the quite familiar questions such as doubling rather than retrieval of number facts from memory.

In exploring further how the children learnt from the first part of the Numeracy lesson it is useful to consider some of the exchanges which took place during or after the lesson. Perhaps some of the most interesting remarks took place in the smaller work groups after the main teaching session. In discussion of fractions, during which the teacher coloured in part of the fraction, (for example, 12/13ths she showed by colouring in 12 of 13 boxes) the following exchange took place:
The teacher asked a problem solving question but answered it with a short and factual answer related to the visual presentation. The larger number could go on top but the total would be over a whole. The researcher asked one of the boys in the following group activity why the larger number did not go on top: the answer was, ‘You daren’t do it the other way round’ It is interesting that this boy saw this as something that he would not dare to do, without any reason given, nor did he appear to find a reason necessary. This draws attention to the subtleties of language, and in this case to the fact that the pupil took the teacher’s words literally.

The relationship between language and numeracy appears to be a complex relationship and the following research (Manor et al., 2000) illustrated some of the difficulties in considering that relationship. They argued that children with developmental language disorder have more difficulty than might be predicted with arithmetic. In a study of 6 year olds and a matched control group, the children with language difficulties had much greater difficulty with reasoning principles of numeracy – the relationships between addition and subtraction and equivalence.

In a study of forty two children with developmental language disorders and a control group of children matched on gender, cognitive abilities and age (the control group
were a little younger) the authors found that the children with language difficulties were poorer on conservation of number, addition and subtraction. Moreover the authors found a correlation between the severity of the language disorder and the difficulties with arithmetic, and that the severity of the difficulty with arithmetic in those children with a language difficulty was greater than would have been predicted from cognitive testing. The authors argued that arithmetical difficulties were associated with but were greater than the language difficulties would have predicted, understanding and expressive language difficulties were associated with number reasoning and arithmetic operations (numerosity).

The authors have attempted to separate the two variables, language and arithmetic, and in doing so have made some surprising assumptions. They argued that arithmetic was not a language based skill and that numeracy was a, 'preliterate skill manifested as early as infancy'

(Manor, Shalev, Joseph and Gross-Tur, 2000, p.71).

It is unclear why they made this assumption, particularly as their assessment of arithmetic skills of children with developmental language disorder appeared to rely heavily on the understanding of language with questions such as:

**TEXT BOX 5.4 Questions put by Manor et al., 2000. p.76**

'How many blocks are there now?'

'Are there more animals than before?' (one animal had just been removed)

'How do you know?'

It may be argued that the children with the language difficulties were placed at an additional disadvantage by such questions. Such children may be more likely to give
what they perceive as an 'expected' answer. Finally, the authors have pointed out that 20% of their experimental group – those with language difficulties, also had attention difficulties. In fact 9 of their sample of 45 had Attention Deficit, Hyperactivity Disorder. It is not quite clear how this was established but it might be argued that this would affect in a negative manner the results of the children in assessment.

The issues discussed by the authors could be applied to the understanding of the children in the numeracy session. Questioning in the numeracy session appeared to depend on firstly the medium of attention, then on shared understanding of the language used, and lastly upon the child's numerical reasoning. However, in the sample of 256 children in the present study, it is probable that a number had difficulties with language and it would not be possible to argue as Manor and colleagues did that the arithmetic in the teachers' questions and in the post-test items was 'not a language based skill' (Manor et al., 2000, p. 75) The post-test items were read out by the researcher to the children and so an understanding of what was being said was necessary to carry out the arithmetical question. Arithmetic could not be satisfactorily separated from the medium of language. However, it appears quite reasonable to suppose, as do these authors, that language and arithmetical difficulties may be associated. It should be noted however that in some cases another factor (for example attention and concentration difficulties) may be responsible for both language and arithmetical difficulties.

5.3 Question types and constructivism

This study has investigated question types and the learning of children in the Numeracy session. However, in carrying out the literature search it became apparent that there were researchers who tended to assume that problem solving
questions somehow evoked what was called 'higher order thinking ' without providing a clear rationale for what constitutes higher order thinking (Galton et al., 1999; Perry et al., 1993; Newstead, 1998; Saxe et al., 1999). On the other hand there were those who emphasised the part played by quick computation, and number facts in memory (Ashcraft, 1992; Butterworth, 1999; Dehaene, 1996). A lack of familiarity with number facts may affect attainment in numeracy; for example, Butterworth (1999) attributed some variance in cross cultural differences in mathematics achievement to practice and speed in this area, particularly to practice in computation.

This tension appears to be between those who espouse constructivism in the classroom and those who emphasise the teaching and practice of computations, and the focus on mental mathematics drill in the Numeracy Strategy. The historical route of mathematics teaching appeared to have swung between these positions and has been discussed in detail, some years before the Numeracy Strategy was devised (Richards, et al., 1983). At this point it is worth considering the concept of constructivist learning in application to numerosity. The constructivist position is made clear by this quotation:

‘every child must build up an appropriate conceptual structure during the first six or eight years of life’ (Richards, Steffe and Glaserfield, 1983, p.113) The authors discussed constructivism relating to the experience of the child rather than something existing independently. This child constructs his or her arithmetical concepts through experience with counting; the authors have studied children in early counting exercises and believe that there is regularity in how children establish early arithmetical concepts.

Both Newstead (1998) and Boaler (1997) pick up a constructivist position. Newstead assumed that what were termed 'reform classes' would have more of the problem
solving questions than the traditional classes. Children given open questions (Galton et al., 1999; DfEE, 1999; Kyraciou, 1995) conceptual questions, (Boaler, 1997; Brown and Wragg, 1993; Perry et al., 1993) were in some way encouraged to construct their own learning. Boaler has compared mathematics in two secondary schools with contrasting approaches, Phoenix Park, and Amber Hill (Boaler, 1997), and has argued that the learning was greater where the pupils were able to construct their own learning, rather than in the traditional school.

A contrasting point of view to the constructivist one might be that familiarity and practice in mathematics, particularly in number facts, was responsible for much of the difference in speed and accuracy in cross cultural studies. Differences in children’s numerosity were shown by the Third International Mathematics and Science Study (Keys et al., 1997). The part played by familiarity with number facts was emphasised by Ashcraft in 1995 and Butterworth in 1999, and is one of the tenets of the Numeracy Strategy (DfEE, 1999). Taking a historical perspective it has been argued that the teaching of arithmetic or computation in the twentieth century has often relied more on drill and practice derived from behaviourist principles (Richards et al., 1983). The authors pointed out that arithmetic teaching treatises borrowed Thorndike’s principles to fix bonds between numbers without appreciating the need to establish awareness of the relationships between the number facts (Thorndike, 1922).

There may be a pendulum swing between the two positions, - the emphasis which Cockroft (1982) placed on the application of mathematics in real life situations, and the emphasis of the Numeracy strategy on familiarity with number facts. It could be argued that, ‘there is the danger that psychological theories will distort the curriculum in application’ (Richards, Steffe and Glaserfield, 1983, p.115).
However, from this study it would not have been possible to group teachers into traditional and reform or alternative (Newstead, 1998, Saxe et al., 1999) which appeared to be a subjective grouping. Teaching group A asked more problem solving questions than teaching group B, and this is a characteristic of reform teachers, but they also asked more computation (drill and practice) questions than group B.

Secondly, it is argued that while it is clear that for practical use children should learn how to apply mathematics in everyday living, the assumption made by some researchers (see Table 2.3 and 2.4) that certain types of questions require higher order thinking is difficult to support and may be an adult construction of what the child does to answer the question. It is argued that the definition of higher order thinking in numeracy would be difficult to agree on; it is difficult to decide why higher order thinking applies to a *how* question and not a *what* question. For example a series of calculations in working out a percentage may place a load on working memory and processing ability, while the question: *how did you do that?* may be relatively simply and quickly answered. However, the first is a computation and rule recall procedural question, while the second is a metacognitive question, requiring the pupil to think about his or her own thought procedures. This would be defined as a problem solving question by most of the authors referred to in Chapter 1 (Tables 1.3 and 1.4). In this study it was not clear how a question demanding a series of arithmetical operations, such as multiplying two digit figures, might demand thinking of a higher or lower order than the one asking the child to reflect on how s/he reached an answer.

However the greatest obstacle to determining whether a question evokes a certain type of thinking is that it is not always possible to know what a child does with a question and this raises again the issue of validity referred to in Chapter 1. Section 4, Chapter 2. Section 4.3, and in Chapter 5. Sections 1.5 and 2.1. Assumptions about a
child's thinking from his or her use of strategies may not be well grounded. Children may continue to use primitive counting types even when they may be able to use more sophisticated ones and it is not always possible to infer from a child’s behaviour what s/he actually is capable of (Richards et al., 1983).

It is argued that higher order thinking may be confused with language competence; if such authors as Jordan, et al., (1994) are correct there may be different processes used depending on the number of the words in the question, and Snorre (2002) has claimed that the same question presented verbally or in number form is differently processed and that the assumption that certain types of question, or thinking are somehow of a higher cognitive order than others, is unfounded. Computation questions may have a series of complex calculations, and require the child to keep them in working memory. Conversely, the question, How did you do that? may evoke the very simple answer I guessed, I just knew. It is suggested that the part played by language may cloud the issues. Perhaps what is viewed as higher order questioning demands a more sophisticated understanding and use of language, and it may not be possible to separate out children’s numerosity from their competence in language.

5.4 Teaching and learning in the Numeracy Strategy for educational psychologists

It is argued that educational psychologists have a research and investigative role in the study of questioning, teaching style and learning in the numeracy session. They have a role of advice, support and advocacy for those children who are not able to access the content because of their level of understanding. Educational psychologists have a role in considering what interventions or recovery programmes
may be suitable for those children and how they might be implemented either within
or outside the framework of the Numeracy strategy.

5.4.1 Educational psychology, teaching and learning

The arena in which educational psychologists work is the school, and although it is
rarely formally stated, they are probably most often consulted about how to enable
children to learn and about any matter which might in some way interfere with
learning. Educational psychologists are psychologists who have chosen to specialise
in the application of psychology to the education of children although it appears to be
the case that, for a variety of reasons, they have become increasingly concerned with
children who have special educational needs. This may have resulted in educational
psychologists being less associated with teaching and learning and is reflected in the
Department of Education and Employment Report on the future of educational
psychology (DfEE, 2000), which, while pointing out that schools want more time from
educational psychologists, did not emphasise learning. However, the working party
report on the future of educational psychology referred to support for children at an
earlier stage of special educational need (DfEE, 2000), to more proactive and project
work and to more training for teachers.

To some extent it is also reflected by the Division of Educational and Child
Psychology (2002). In reviewing the strategic plan from 2002 to 2005, it refers to
developments in dyslexia, and in Assessment and Intervention but does not discuss
other areas of learning such as mathematics. Other areas of school life appear in
educational psychology journals such as *Educational and Child Psychology* or
*Educational Psychology in Practice.* - children's behaviour, self esteem, friendships,
reading and spelling skills journals. There are fewer papers on cognitive psychology
and its application to learning, especially the acquisition of mathematical skills
Numeracy has been traditionally less an area for investigation by educational psychologists than has literacy. There are exceptions in that educational psychologists may have interests in dyscalculia, for example, Lawson gave an account of presentations at the International Congress of Psychology (Division of Educational and Child Psychology newsletter, March 2001), but searches of databases indicate that such papers are more likely to be found in journals dedicated to mathematics teaching such as Mathematics Research in Education (ERIC, PSYCHLIT and SOSIG databases, 1997-2002). It has been necessary to go to either journals focussing on cognitive psychology or journals of mathematics research for material.

How children learn in the whole class context of the numeracy lesson demands research skills, to evaluate the teaching and learning in the numeracy session, and the learning in recovery programmes (Dowker, 2001, Steffe, 1992), and an interest in developments in cognitive psychology with particular application to mathematics. For those with difficulties educational psychology has the remit of advice and intervention and this is discussed below. It is argued that the area of screening for difficulties in mathematics should be part of the repertoire of the assessment of the educational psychologist. Performance in Standard Attainment tasks will be helpful in initial identification of those pupils with mathematical difficulty but in order to decide on interventions which will be directed to the next zone of understanding it will be important to have as full a picture as possible of the child’s difficulties, particularly in the light of the possible relationship of arithmetical difficulties with language development (Snorre, 2002).
However, educational psychologists are well placed by their training and involvement in schools to consider the implementation and effects of Government policy such as the Numeracy Strategy and it is argued that they would be well placed to evaluate the effects of initiatives on learning and to work with teaching staff in helping them to evaluate their own teaching. The study above identified different questioning styles which might form a framework for teachers embarking on their own professional development. The use of video recording could be developed as part of the educational psychology service contribution to teacher professional development and monitoring duties of head teachers. Head teachers may wish to use of video feedback in professional development; to ensure a positive and objective learning experience for teaching staff, this might be used in conjunction with either the educational psychology service or Higher Education Centres specialising in teacher training. Teaching staff engaged in self evaluation may wish video feedback and post-test performance to allow them to gauge the learning of individual children, and the class as a whole, from the Numeracy session.

5.4.2 Children experiencing difficulties in mathematics and educational psychology

While it is argued that educational psychologists have a role in teaching and the learning of all children, their present role is more likely to be involvement with that sizeable number identified by this study who, in any class, may be experiencing difficulties with the Numeracy strategy. This number may not reflect the number of children who are on the register for children with special educational needs, but may be considerably greater. The educational psychologist will want to consider what the learning is for the child who has significant difficulties in the mathematics lesson and how inclusion is implemented.
The study above has indicated that the whole class teaching session may be too difficult for as many as ten per cent of the class, that is, that the concepts being taught are not within the zone of proximal development for those children (Vygotsky, 1978). It has indicated that there may be an additional challenge in problem solving for those children with language difficulties. Policies at present emphasise inclusive education and advice has been provided on how to achieve this, for example with pupils who have a range of learning difficulties (QCA, 2001). However, while such advice sets out what most pupils with learning difficulties might achieve, for example counting to 10, the issue of how to include these pupils in the whole class teaching session, and introduction of much more complex material, is not dealt with. There appears to be a tension between the demands of a Vygotskian constructivist approach to learning numeracy, starting with the learning of the child, and questioning to provide scaffolding from that point (Wood, Bruner and Ross, 1976; Wood, 1991) and the demands that the child must experience the whole class session (Vygotsky, 1978).

There appears to have been little investigation of the effects of the whole class teaching of the Numeracy session on children with mathematical difficulties. If Newstead (1998), discussed in Chapter 1, Section 4) is correct and there is a relationship between mathematics anxiety and difficulties with mathematics it is possible that the experience of children who do not understand the questioning, is one of anxiety at worst and non involvement at best. We do not know what the further public element of questioning in a whole class situation may add; once again this appears to be an area for investigation by the educational psychology services. Such investigation and evaluation will allow informed recommendations to be made for all children, and for the children who may be particularly affected by the issues.
Chapter 6

Evaluation

6.0 Evaluation of the study

Chapter 6 will evaluate the methodology of the study, highlight the limitations and consider some of the more general issues surrounding the subject.

6.1 Control of extraneous variables. How far is it possible to generalise from this study?

This study attempted to apply some of the principles of cognitive psychology to learning, in considering what type of questions teachers ask, and how children learn to do mathematics, particularly in the context of the whole class teaching required by the Numeracy Strategy. In applying research methodology to real life situations there are variables which cannot easily be controlled. For example, the learning of the children in each numeracy session may have depended on many variables other than the teaching; it is likely that the number of children with mathematical, language or attention difficulties varied in each class; some children may have had previous experience of the subject either at school or at home; some teachers chose topics which may have been intrinsically more difficult to grasp, or pitched their presentation in a more complex way; others may have been revising a topic learned previously.

A variable central to the children’s attainment in the post-test, the measure of learning, and which was not controlled, was lesson content and difficulty. In this study the areas covered included multiplication of 2 digit numbers, the degrees in a straight line and in right angles, the making of a frequency chart for likes and dislikes of food, a Carroll diagram, and different strategies for adding large numbers. In the
case of this study it was difficult to ascertain the content of the lesson beforehand; nine of the ten teachers when asked were not sure which subject they would be covering on the date selected. As the post-test had to reflect the content of the lesson, the teaching for one class used calculations with larger numbers and so it is not possible to say whether that classes' poorer performance was due to the intrinsically larger demands made by the questions or because a lower rate of success during instruction led to a lower level of engagement. Nor does it appear possible to control for children's previous experience of that subject. A pre-test, the selection of a larger sample and the resources to carry out the observations during a shorter time span would to some extent overcome the effects of such variables.

A more satisfactory experimental situation would provide for a pre-test, control the level of difficulty of both teaching material and of post-test questions. With the prescriptive nature of the Numeracy Strategy (DfEE, 1999) it would be possible to achieve a greater degree of similarity in content; for example the researcher could plan observations over a shorter period and request the observation lesson be a topic such as multiplication of two digit numbers, or a topic generally presented during that half of the term. However, it is likely to remain difficult to control for complexity of the lesson, and the children's previous experience of the subject. Using the example of multiplication of numbers greater than 10 ('long' multiplication), in the same age group one teacher chose to concentrate on the multiplication of smaller numbers and present only one strategy, while another selected larger numbers, of two and three digits, and taught the use of two strategies.

In view of the difficulty of controlling these variables a more satisfactory method of assessing the learning of the children would be to present them with a pre and post-test. This would allow an assessment of the knowledge of the children in the
particular area to be taught, and would provide an indication of children with mathematical difficulties.

Comparison of a pre-test with a post-test score would have more accurately reflected learning during each lesson, or even shown an increase in error rates, (Boaler, 1997). Without a pre-test it is not possible to know to what extent performance on the post-test was attributable to the learning in the lesson or to previous learning, either in or out of school. However, the practicalities of setting up the observation worked against this. The researcher had to reassure the head teachers that the usual timetable would not be disrupted, teachers were unsure what the lesson content would be when arrangements were made, sometimes up to 12 weeks in advance. While it is felt that both those practicalities, particularly lesson content could be overcome (see above), other psychological variables should be taken into account. The first is the necessity of asking children to undergo two formal tests. Such tests in which children are not allowed to confer or in the case of those with learning difficulties to ask for help from a teaching assistant, are rare in most primary schools in this area. Secondly, if teachers were present during the pre-test it would be difficult for them to avoid teaching to those items during the lesson; there are additional practical difficulties in asking them to leave the classroom. It is acknowledged that a pre- and post-test would give a more accurate indication of learning, and the above practicalities could be overcome, but there would be a considerable advantage in having a teacher who was not the class teacher available, and in giving thought to the ethical issue of how to avoid stress for the children, particularly those with learning difficulties.

In this study the carrying out of the post-test was one of the more difficult parts of the operation and the part most likely to cause most ethical difficulties; in general all children in a class are not asked the same questions in a formal written test, apart
from the Standard Attainment Tests (SATs), and it was noted by the researcher in one class that there was some consternation on the part of the children who were not able to answer the questions and were not allowed to ask for help from friends. Efforts were made by the researcher to ensure that all tests finished with a very easy question, but it should be remembered that in each class there were children who scored zero, and it is hard to justify the increased testing resulting from a pre- and post-test for those children with mathematical difficulties. However, the Numeracy Strategy has reintroduced a period of teaching of the whole class, with a prescriptive curriculum and at the same time a renewed commitment to inclusion may have resulted in more children with learning difficulties attempting to follow teacher questioning geared to other abilities and understanding.

An investigation of whether there are relationships between teaching style and children’s learning could be more satisfactorily carried out by the use of a larger and balanced sample of schools. It is likely that the researcher will find more similarity in the structure and content of mathematics lessons now than before the Numeracy Strategy was introduced.

From this sample, it may be argued that educational psychologist could make greater use of video recordings of lessons, working closely with teachers. The two teaching types may be markers on a continuum of teaching and questioning styles, and serve as a framework for teachers to consider their own interaction styles. For example, teachers may switch from one style to the other depending on subject content and the demands of the class. In this case the teacher would be adapting the nature of the interaction to the context, and this has been claimed to be the art of pedagogy (Galton et al., 1999).
6.2 The Validity of the Classification System.

Chapter 2, Section 4.3, Chapter 5, Section 2.1, and Chapter 5, Section 3, have raised the difficulty of determining the validity of the classification system. It could be claimed that this classification had face validity as judges had a high level of agreement on question types and could explain their classifications easily. For example, they pointed out that problem solving questions were longer than computation questions. However, establishing construct validity is difficult where there is no easily obtained independent criterion against which to measure the classification, for example in the case of the classification of mental illness and certain physical illnesses (Gross, 1992). Studies may admit that the classification is subjective (Newstead, 1998) or bypass the issue (Perry et al., 1993; Boaler, 1997; Galton et al., 1999).

In this study there are two constructs against which the classification might be measured. The first is the teacher's intention in asking the question, (Morgan and Saxton, 1994) and the second is what the child has done to answer the question. The ORACLE study (Galton et al., 1999) although not stating this explicitly, chose as the construct the teacher's intention, because observers classified an open question as one where the teacher actively sought more than one answer. However, the teachers were not asked their intention; different observers classified question types, issues of reliability and validity. In the case of the present study, teachers could have classified a sample of their own questions, providing a validation of the system, but this was not built in to the design.
The second construct, possibly more relevant to the learning of the child, is what s/he does to answer the question, which may not match the teacher's intention. (An example is the question and answer presented in Chapter 5, Section 1.3, Text Box 5.1). The child might be asked how s/he arrived at the answer, (Siegler, 1987a). Such an approach would necessitate individual assessment, carried out in Maths Recovery schemes to work out programmes for each child (Dowker, 2001; Steffe, 1992). One immediate advantage would be the emergence of a clearer picture of how the child addresses the problem. To avoid the weaknesses of self reporting (for example, communication difficulties) there are other experimental designs, such as requiring the child to share cognitive capacity with another task, giving an indirect indication of whether children are retrieving information or working out a calculation (see McKenzie et al., 2003). Following the judges' discussion about what children did to reach an answer, this study asked children to show the steps taken in some questions to throw light on their processing.

In conclusion, it is acknowledged that if questions are categorised to reflect teacher intention as the underlying construct, the classification would gain validity by asking the teachers to classify their own questions. However, if children's learning is the construct against which the question types are measured, additional information is needed about how the child has answered the question. Such information would allow firmer conclusions to be drawn about how the child's learning is related to the teacher's question type.

6.3 Video Methodology

In the evaluation of this study, there are benefits and costs in the use of video recording. However, it is argued that in the interest of as full a record as possible the use of video recording was justified. 'Completeness' is one of the criteria put forward by Simco and Warin (1997) in assessing the trustworthiness of the research
outcome. It might be argued that the record was a much fuller one than could have been gathered by an observer, or even an observer with a tape recorder, and that there is an additional advantage in having one person recording because of consistency of method and interpersonal variables. The study carried out by Perry and colleagues (Perry et al., 1993) illustrated the risks to reliability in their cross cultural research in the use of 30 observers using no kinds of recordings, in three different countries and the conversion of their observations into summaries.

The use of video recording allowed viewing and reviewing of the material, but is likely to have carried certain costs. A disadvantage is the effect of the presence of a VCR on the teacher and class,

‘there is no doubt that observation is the most intrusive of all techniques for gathering data’ (Simco and Warins, 1997, p.55).

In all cases, the teachers had prepared the pupils for the presence of the video recorder and asked them not to look at it but to pretend that the lesson was normal. One teacher was however noticeably nervous at the start of the lesson, and 7 of the 10 made a remark to the effect that they were relieved when it was over. Three teachers appeared anxious and ill at ease despite the attempts of the researcher to be as unobtrusive and reassuring as possible. The preparation for the video recording in the setting of suitable dates and the requesting of head teacher and parental permission meant that it was difficult to keep the occasion low key and it appeared that this effect was larger in schools where the researcher was unknown to staff.

Because of this it is difficult to claim that the teaching behaviour and the children’s behaviour was typical of a usual numeracy lesson. There should therefore be
reservations about how far it is possible to generalise from these results to the
general practice of the teaching population. However, it may also be argued, as do
Simpson and Tuson, (1995) that many experimental methods change the behaviour
of the participants. They have pointed out that even the act of filling in a
questionnaire can change the views of participants. The effects of the video
recording may have been reduced by allowing staff to become accustomed to the
presence of the observer with the video recorder. Over time, in video recording
student teachers over their teaching practice there appeared to be little effect on their
teaching (Simco and Warins, 1997). If several observations of each teacher had
been carried out more accurate data may have emerged.

The issue of whether teachers would change their questioning because of their
knowledge of the observer’s purpose should also be considered. It was not possible
nor ethically desirable to have experimental naïveté. The researcher had considered
the issue of informed consent when the heads and teachers were approached; in
keeping with ethical guidelines (BPS, 2000) the researcher informed heads and
teachers that the area of research was questioning in the numeracy session. The
conclusion of the ORACLE study which pointed to teachers’ decline in
professionalism and of their practice of talking at the class (Galton et al.,1999) was
not borne out in this study but this may be because teachers changed their behaviour
during observation, and the ORACLE methods of time sampling may have resulted in
fewer changes.

The use and interpretations of recordings is part of a larger debate about the
objectivity of the observer. Simco and Warins (1997) have raised the lack of
objectivity in the choice of whom and when to record, how full a description to
transcribe, and how to interpret the transcription. Their solution was for the
researcher to promote his or her own reflexivity and to challenge his or her interpretations. Mead described reflexivity as, 'the turning back of the experience upon [him or herself]' (Mead, 1934, p.134) and Delamont as, 'a social scientific variety of self consciousness' (Delamont, 1991, p.8) In contrast to this was the view offered by Maxwell in discussing the validity of qualitative research: ‘As observers and interpreters of the world we are intrinsically part of it; we cannot step outside our own experience to obtain some observer independent account of what we experience’ (Maxwell, 1992, p. 66)

This issue of the nature of objectivity has been picked up in relation to inter-judge agreement; it has been argued that such agreement is less rich than competing interpretations and that inter-judge agreements are an inevitable consequence of a power structure (Simco and Warins, 1997). In the pilot study judges appeared to be quite happy to voice and support their own case for classification, and there was no evidence that judges changed ratings in accordance with an implicit power structure.

6.3 Future research possibilities

The cluster analysis indicating two different teaching types could be applied to a larger sample with either a greater degree of control of lesson content and the proportion of children with special educational needs, or with a design to investigate the effects of make up of the class. It would be interesting to know what teachers see as the most challenging questions in their own transcripts, and whether they had certain pupils in mind when they asked these. It does appear that teachers and head teachers are interested in video feedback for a range of purposes, professional development being one of these. It may be that a framework could be developed so that individual teachers could place themselves on different continua; there may be
two or more questioning types, and teachers who fall into traditional and reform types of teaching (Newstead, 1998; Saxe et al., 1999). The relationship between class management and computation was a further unexpected finding which would be worth further investigation; these types of questions may further describe teaching style, and the possible effects on learning.

One of the findings which stood out was the degree of difficulty a sizeable proportion of the children had with mathematics and the number in each class who did not appear to have grasped the topic or be able to answer very simple questions on it. The children's confusion in the use of 'rule following behaviour' in the learning of mathematics was perhaps an indicator that teaching a variety of strategies may not be useful for all of the class although it may encourage the development of numerosity (Nunes, 1996; Butterworth, 1999) for some others. One useful direction would be to compare learning of different groups given one or more than one strategy, and comparing teaching through different ways of questioning. Boaler's finding that more was remembered of projects led by the pupils in secondary school may be reflected in primary (Boaler, 1997).

From this finding arise the issues of differentiation, of teaching method, and of anxiety and failure. This study indicates that the issue of differentiation in the numeracy session and of satisfactory inclusion should be further investigated. In the light of the number of children who scored zero in the post-test it appears possible that some children do not understand the majority of the questioning and teaching. While inclusive practices have emerged from children's rights not to be segregated, it may be argued that children also have the right to have presented to them appropriate material geared to their zone of proximal development (Vygotsky, 1978). It appears particularly important to assess the learning of all children in a whole class teaching situation as greater emphasis is now placed on minimizing removal of
children for lessons such as Mathematics Recovery schemes (Steffe, 1992, Dowker, 2001). Given that Newstead (1998) found a relationship between ability in mathematics and mathematics anxiety it might be argued that this was an ideal opportunity to look at the feelings of much younger children towards mathematics and to investigate the variable of social or public anxiety – answering questions in the whole class session (Newstead, 1998), particularly for those children who have difficulties with both the language of the questioning and arithmetical processing.
References


Fraser, B. (1989). *Assessing and improving classroom environment*. Key Centre for
School Science and Mathematics, Curtin, Australia


processing differences in young and elderly adults. *Psychology and Ageing, 6*, 474-
483.


attainment of boys and girls at school. *British Educational Research Journal 27*(2),
125-140.


*Psychological Review, 79*, 329-343

London: Hodder and Stoughton

Education (Bedford Way Papers, no. 34).

*J. Exp. Psychology: General, 123*(3), 284-296.

Mathematics, 12*(1), 8-14.


Oxford Conference. Convener: Dr A Dowker, Univ. of Oxford, Department of Psychology, Oxford.


Appendix 1 Part of a transcript

t teacher, b: boy, g: girl, initial: name

t Now can we do pairs of numbers to 100?
so if I give you the number can you give me its twin?
25?
g 75

t 45?
b (T) 65
t 65? Careful - -J
b 35
T 25? C
g 25
t no 25?L?
b 75
t Good boy. OK 15?M?
b 85
t Good boy, 85, 5?M?
b 95
t Excellent M well done
65?B come on quick quick
g ??
t A?
b 35
t Ok 55? Y?
g ? 55
nt not 55 - -?
g 45
t Right now can we do similarly pairs of number to 1000?
I'll start you off
t 750 - and ?A?
b 250?
t 450T?
B 650
t is that quite right? Have another go T?
b 550
t Yes right 350?350? C?
g 650
t 150? 150? L?
g 850 -
t yes 50 come on quick M
b 950 -
t excellent one last one?350 A?
g 650
t- excellent OK
Appendix 2  A sample of questions asked in the post-test

Lesson content: dividing and remainders; rounding up and rounding down

Computation

- 13 divided by 3  
- 22 divided by 5  
- 36 divided by 4  
- 22 divide by 3

Problem solving

- I have 23 fish fingers and 5 children How many will each get?  
- We need to take 23 children on a trip and we can fit 3 in each car  
- How many cars? PS  
- a pencil costs 15 p. How many can you get for 45p?

Lesson content: relationship of multiplication to addition: multiples of 5 multiples of 10: halving doubling: multiplying two digit numbers in different ways – decomposing and adding, doubling using multiples of 10

Computation

- 10 times 10  
- 5 times 8  
- 4 times 5

Problem solving

- How will you tackle 4 times 13  
- How will you do 8 times 17  
- How could you multiply 8 by 20?  
- How can you check your answers in division?

Lesson content: Number bonds to 100. Number bonds to 1000 Money up to a pound

Computation

- 25 add 75?  
- 65 add? equals 100  
- 35 add? equals 100  
- 100 take away 71 ?

Problem solving

- If I have a pound and buy a Mars Bar costing 29 p how much change will I have?  
- I want 4 rubbers and each one is 10 p. How much change would I get from a pound?  
- How could you work this out in two different ways?  
- How do you take away 245 from 1000?
Appendix 3 Graphs showing results of post-test in the 10 classes
Those with plain outer background are Type A teaching style, those with buff outer background are Type B, and the blue inner frame shows those where there are significant differences between computation and problem solving.
University College London

Doctoral Programme for Practising Educational Psychologists (DEdPsy)

Professional Practice Assignments

Submitted in part fulfillment of the requirements for the Continuing Professional Development Doctorate in Educational Psychology (DEdPsy)

Name of Course Member: Deborah Mary King
# Contents

<table>
<thead>
<tr>
<th>Assignment one</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline assessment and the identification of</td>
<td>3</td>
</tr>
<tr>
<td>special educational needs: a project in an Education Action Zone</td>
<td></td>
</tr>
</tbody>
</table>

| Assignment two                                     | 49   |
| Measuring the outcomes of consultation: part of a best value approach. |      |

| Assignment three                                    | 93   |
| The Numeracy Strategy or recovery programmes for pupils with difficulties in numeracy. |      |

| Assignment four                                     | 132  |
| Communication skills in the statutory process: the professional development of one educational psychologist. |  
Professional Practice Assignment

Submitted in part fulfillment of the requirements for the Continuing Professional Development Doctorate in Educational Psychology (DEdPsy)

Name of Course Member: Deborah King

Title of Assignment:

Baseline assessment and the identification of special educational needs: a project in an Education Action Zone

Assignment number 1

Core Curriculum area to which this assignment topic relates:

Assessment and Intervention
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0 Abstract</td>
<td>5</td>
</tr>
<tr>
<td>2.0 Baseline assessment consultation in an Education Action Zone</td>
<td>7</td>
</tr>
<tr>
<td>2.1 Baseline assessment and consultation procedures</td>
<td>8</td>
</tr>
<tr>
<td>2.2 Implementation and feedback</td>
<td>10</td>
</tr>
<tr>
<td>2.3 Baseline assessment data in the Education Action Zone and in the 10 National Sample, 1999-2000</td>
<td>11</td>
</tr>
<tr>
<td>2.3.1 Baseline assessment scores in the National Sample and the EAZ</td>
<td>12</td>
</tr>
<tr>
<td>2.3.2 Profiles of children with low scores</td>
<td>13</td>
</tr>
<tr>
<td>2.3.3 Comparability of data in the EAZ and the National Sample</td>
<td>13</td>
</tr>
<tr>
<td>2.3.4 Implications for provision</td>
<td>14</td>
</tr>
<tr>
<td>3.0 A review of selected theory and research relating to baseline assessment</td>
<td>16</td>
</tr>
<tr>
<td>3.1 Technical issues and baseline assessment</td>
<td>17</td>
</tr>
<tr>
<td>3.2 Validity for children of different cultures</td>
<td>22</td>
</tr>
<tr>
<td>3.3 Baseline Assessment in screening for children with special educational needs</td>
<td>23</td>
</tr>
<tr>
<td>4.0 The use of baseline in screening and the criteria of Wolfendale and of Lindsay and Desforges</td>
<td>31</td>
</tr>
<tr>
<td>4.1 Parental involvement</td>
<td>32</td>
</tr>
<tr>
<td>4.2 Baseline data in the Education Action Zone</td>
<td>38</td>
</tr>
<tr>
<td>5.0 Conclusions and recommendations</td>
<td>40</td>
</tr>
</tbody>
</table>

References

Appendices:

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0 Baseline assessment scales</td>
<td>46</td>
</tr>
<tr>
<td>2.0 Developed model of baseline assessment (Lindsay and Desforges 1998)</td>
<td>47</td>
</tr>
<tr>
<td>3.0 Wolfendale's questions (Wolfendale 1993)</td>
<td>48</td>
</tr>
</tbody>
</table>
1.0 Abstract

This assignment describes a small scale project to improve early identification of special educational needs within an education action zone. It considers the use of baseline assessment as a first school based screening for special educational needs; educational psychologists met with reception teachers and special needs coordinators following baseline assessment, to agree on suitable strategies for children whose profile of scores caused concern. The following term the progress of those children was reviewed to decide whether the strategies should continue and appropriate placing on the register for children with special educational needs. At this meeting other reception children were discussed who had not been originally identified by baseline assessment. An unplanned part of the project was the scrutiny of the baseline scores across the education action zone to see if common areas of need were apparent.

In section three views on baseline assessment in 1993 have been selected and their application to this small scale project are considered; Wolfendale voiced criteria which baseline assessment should satisfy (Wolfendale, 1993) and Drummond had considerable reservations about the use and ethos of baseline assessment (Drummond, 1993). Technical aspects which may affect the use of baseline assessment as a screening instrument are also discussed in the light of more recent consideration of baseline assessment by Lindsay and Desforges (Lindsay and Desforges, 1998) and a survey of users of baseline (Lindsay et al, 2000). The final section evaluates this project in the light of these views. The advantages and disadvantages of consultation following baseline assessment in this specific
Education Action Zone project are discussed, with reference to the criteria proposed by Wolfendale (1993) and Lindsay and Desforges (1998), and some implications for the practice of educational psychologists are considered.

Throughout the assignment the baseline assessment referred to is the Schools Curriculum and Assessment Authority version in Appendix 1 (SCAA, 1997a) but it is recognised that other approved schemes are in use.
2.0 **Baseline consultation in the Education Action Zone**

Education Action Zones were set up by the current government in the 1998 School Standards and Framework Act (DfEE, 1998) with the aim of improving educational attainment in socially disadvantaged areas. Funding and support was intended to come from a mixture of private, voluntary and government sources. To determine suitable areas various indicators were used, mainly relating to socioeconomic status, such as free school meal entitlement which appeared to be related to performance of children on baseline reading, maths and rhyme scores (Tymms et al, 1999). One group of schools successful in its bid to become an EAZ was centred on the south eastern area of Oxford City and incorporated eight first schools. In view of the present thrust of government policy towards early intervention and preventive work (DfEE Green Paper 1997) the educational psychology service decided that baseline assessment should be used in identifying children with special educational needs, and establishing strategies and interventions. It was suggested to the Director of the Education Action Zone (EAZ) that extra consultation time with educational psychologists be provided around baseline assessment for first schools in the Education Action Zone, the aim being to identify as early as possible within their school career, those children with special educational needs, and to put in place early interventions or strategies. Subsequently a model of the process was drawn up, based on the model of Lindsay and Desforges (1998) (Appendix 2).
2.1 Baseline assessment and consultation procedures

Baseline assessment is the first assessment after the child starts school and must be carried out within the first seven weeks. It became mandatory in 1998 (Department for Education and Employment, 1998) although assessments had been used in some authorities much earlier (Wolfendale, 1993). Appendix 1 shows the scheme drawn up by the Schools Curriculum Assessment (SCAA, 1997a) which was the one used in the Education Action Zone although over ninety different baseline assessment schemes had been approved (Lindsay et al, 2000). The five sections in baseline assessment relate to literacy, maths, language, physical skills, and personal and social development. Lindsay and Desforges (1998) point out that baseline assessment may be used for two purposes; it can be used to calculate 'value added' by the school by comparing the scores of a cohort of children with later scores on Standard Attainment Tasks, and it can be used as an initial screening for Special Educational Needs. The use considered in this assignment was as an initial screening for Special Educational Needs.

A trial had been carried out with a first school outside the education action zone. Following inspection, one area for action had been the early identification of pupils with special educational needs; the registered inspector had suggested the placing of children on stage one of the Code of Practice for children with Special Educational Needs if they scored below 12 on baseline assessment. As the educational psychologist had been the named person responsible for setting up and monitoring action on this, she drew up an approach to the early identification of special educational needs using both the profile of scores, and overall scores from baseline assessment, in addition to information from preschool providers, and from parents.
In the Education Action Zone the two educational psychologists for seven of the eight First Schools decided to adopt this approach. Baseline assessment screening meetings took place in the Autumn term 1999, the Spring, Summer, and Autumn terms 2000, following these procedures and criteria. The bullet points below summarise the steps:

- Class teachers, educational psychologists and special needs coordinators met for additional consultation with staff after baseline assessment, between 7 and 8 weeks after the new intake (baseline assessment is carried out within the first 7 weeks)
- Pupils were discussed in the light of their baseline scores, the observations of the reception teacher and information from parental and preschool sources. The criteria are discussed below. Other children identified by the class teacher were also discussed.
- Strategies were agreed on at that consultation. High scores were also discussed to ensure a sufficiently challenging curriculum. It was decided that the more formal intervention of an individual education plan with targets should be left until review in the following term.
- The reception teacher discussed the strategies with parents when baseline assessment was reported, and received parent input. A low key approach was recommended with as much parent involvement as possible. Some children were placed on the register for children with special educational needs, but more frequently they were monitored without placing on the register.
- The following term it was decided whether these strategies had been sufficient or whether the child should be placed formally on the register for children with special educational needs.
The criteria used in selecting which pupils should be discussed and strategies agreed were either an overall low score, or low scores in sections of the assessment. Children with a score of 10 or less on baseline assessment were considered to be 'of concern'. Although the initial suggestion from Her Majesty's Inspectorate in the trial school had been 12, taking into account the socioeconomic status of the schools' catchment areas and the varied preschool provision, the psychologists decided to drop the score to 10. Profile of scores was also used as a criteria; a score of 1 or 0 in any sections, and particularly a combination of low scores would be a trigger for consultation and strategies. In addition, children who were already on the Code of practice for Children with Special Educational Needs (Department for Education and Employment, 1994) and those highlighted by the teacher or preschool provision were also discussed.

In view of the language content and possible culture bias these criteria were not used for children with English as an additional language but teachers from the Ethnic Minority Achievement Group were informed of low scores (see section 3.2).

2.2 Implementation and feedback

Baseline consultation was implemented in 7 of the 8 EAZ first schools following the intakes for September 2000 and January 2001; lack of psychologist time prevented it being implemented in the eighth. In one of the seven schools screening was not carried out in term two, as two children who had started in term one had emotional and behavioural difficulties of a level to warrant statutory assessment for specialist provision and the psychologist's time was used for pastoral support planning and review meetings.
Informal feedback has come from the special needs teachers and reception teachers, the educational psychologists and the head teachers. Formal procedures for gathering feedback are planned for the coming year. Positive responses have been that the baseline consultation raised awareness of special educational needs and early interventions; negative responses queried whether baseline assessment accurately identified all children with special educational needs, on possible impact on parents, and on extra meeting time required for staff and educational psychologists.

2.3 Baseline assessment data in the Education Action Zone and the National Sample data, 1999-2000

Analysis of baseline assessment data had not been planned as part of this project but, during consultation, it became apparent that little LEA analysis of baseline scores was available. The director of the action zone and the head teachers asked questions about the scores which led to the author scrutinizing the EAZ data and data available from the National Sample, 1999-2000 (Qualifications and Curriculum Authority, 2000). Because of time and limitations on the comparability of the data this has been limited to inspection of means and consequently no firm conclusions may be drawn from such inspection.

Raw data was passed to the author by the LEA, consisting of scores for each pupil in the county for each part of the baseline assessment and national data was gained from the Qualifications and Curriculum Authority (QCA,2000) on a sample of 19277 children who were assessed using schemes incorporating the QCA Baseline Assessment Scales (SCAA, 1997,a) during the year, 1999-2000. These data provided the percentage of children attaining each item in the scales and frequency data for overall scores; they also provided mean scores categorised by gender and
month of birth and some of these data could be compared to the data for the education action zone representing the baseline assessments for 1999-2000. The numbers of baseline assessments done during that year in the EAZ totalled 414 children. However, there were limitations on how far the data could be compared and these are discussed in section 2.3.3 below.

The maximum score in baseline assessment is 30. In the education action zone, out of 414 children over that period, 90 scored 10 or less. However, this figure disguises considerable variation among the schools, the number of low scores in any one intake varying from two to eleven (intakes varied from 26 to 11). Means for the national sample are for pupils between 5.0 and 5.3 as this is the age of children in the EAZ when they are assessed (QCA,2000). Variation in scores was high, with a standard deviation of over one third of the mean.

2.3.1 Baseline assessment scores in the National Sample and the EAZ

Graph 1 shows the aggregate (i.e. total score for each pupil) mean scores of the EAZ schools beside national means and indicates means were lower in speaking and listening, Reading A, Maths A and Personal and Social Development but higher in other areas such as Maths B and Writing. This graph shows all EAZ schools and disguises variation between schools (as do the national means). This is surprising considering the generally low SATs scores for the EAZ schools and may be partly due to assessment procedures which are considered in section 4. Table 1 and Graph 2 show the mean scores for each school in the education action zone. It can be seen that two schools, B and F, have markedly lower mean scores than the other schools and than the national sample. Schools A and G have higher means than the other schools and than the national sample. Once again this was surprising and in the case
of school A further investigation showed some anomalies in carrying out assessment (see discussion).

Table 1

Means of total scores

<table>
<thead>
<tr>
<th>National sample</th>
<th>EAZ Schools</th>
</tr>
</thead>
<tbody>
<tr>
<td>QCA, 2000</td>
<td>A</td>
</tr>
<tr>
<td>16</td>
<td>17.6</td>
</tr>
</tbody>
</table>

Standard deviation for all EAZ schools is 6.29. SD for national sample unavailable.

2.3.2 Profiles of children with low scores

Graph 3 groups the scores to show in more detail the profiles of different groups of low scores. It is useful to look more closely at the low scores as these children were the subject of consultation, and it can be seen that children who score under 10 show particular difficulty in speaking and listening. (Unfortunately it is not possible to extract this information from the national data for children over 5 so we do not know if this is a feature of the national sample). All eight schools showed a relatively large number of children with low scores in Speaking and Listening and two schools, B and F, stand out in the numbers of children scoring 0 in Personal and Social Development. These two schools also show lower mean scores than the other EAZ schools in all areas of baseline, and greater numbers of children with scores at the low end of the spectrum.

2.3.3 Comparability of data in the EAZ and the National Sample

Statistical information was not easily available for the education action zone schools and analyses of the EAZ pupil scores were constrained both by local education authority data, and by national data sample. Inspection of means indicated apparent
differences but too little is known about the national data to compare these statistically with local data for the following reasons.

Firstly, and most importantly for this study, scores have been shown to increase with age (QCA 2000, Hedger and Jesson 2000). As the children in the EAZ were all over 5 years at baseline it was necessary to compare EAZ means only with those from the national sample who were over 5. Most of the national sample data does not separate out scores by age (QCA, 2000, pages 21-23) so that we cannot tell from the national sample what percentage of children between 5 and 5 years 3 months had overall scores of less than 10. Secondly, neither the LEA nor the national sample gave scores separately for children with English as an additional language. Thirdly, analyses of baseline assessment scores carried out by the local education authority did not analyse the separate components of the scores, nor did it separate out the scores for the EAZ schools from the rest of the county. Finally, differences in assessment practice between schools, local moderation procedures and training of new staff raised questions about the reliability of local data.

2.3.4 Implications for provision

Initial consultation and analysis of the baseline scores identified two main areas in most of the school as targets for consultation and intervention. These were communications skills, and personal and social development.

In view of the needs identified in language and communication one initial focus of the Education Action Zone Director was on speech and language; two speech therapists were employed to set up a programme focusing on speech and language in the nurseries. The speech and language therapists will be starting a rolling programme of training for all nursery staff in an approach aimed at increasing and facilitating
children's language. Additionally the EAZ funded a speech therapist for one half day weekly to visit nurseries with the aim of training staff for individual and small group work. The educational psychologists are in the process of following this up with the introduction of the Teaching Talking programme for the early years at school (Locke and Beech, 1991).

It is probable that difficulties in communication skills affected scores in other areas, particularly in personal and social development (Appendix1 section 5). Teachers were most interested in, and felt they had least expertise in, strategies for children who scored low in the personal and social development section. Strategies discussed were inadequate for a small but significant proportion of children who moved quickly towards statutory assessment. In two of the seven schools there were permanent exclusions within the first two terms and both of these caused considerable concern and distress for children, parents and staff. The children had been identified in the nursery and placed on stage three of the Code of Practice; their scores on the Personal and Social Development Items were 0 or 1, and their overall scores were below 8 so that on both criteria above the children would have been identified by baseline assessment, had they not been raised by staff at their preschool provision.
3.0 A review of selected theory and research relating to baseline assessment

In the following section some of the literature is discussed relating to the ethos and effectiveness of early assessment and intervention and its relevance to an Education Action Zone (EAZ). Technical issues to do with test construction (Lindsay and Desforges 1998) and arguments for and against the use of baseline assessment in screening for children with special educational needs are considered in the following section (Wolfendale, 1993, Drummond, 1993, Lindsay et al, 2000). In particular it will consider Wolfendale’s caveats in 1993 when reporting for OMEP, (Organization Mondiale pour l'Education Prescolaire UK, 1993) on baseline assessment, as well as more recent discussion on baseline assessment and special needs identification, such as Lindsay and Desforges (1998) and Tymms (1999). The recent evaluation of the various baseline schemes by Lindsay, Lewis and Philips (Lindsay et al 2000) will be considered in the literature review and in the discussion section.

Lindsay and Desforges (1998) categorised the purposes of baseline assessment into those with a child focus and those with a school focus. Those with a child focus are screening for special educational needs and identification of what need or difficulty exists, setting of targets and strategies to achieve these, and monitoring. This assignment will focus on the first of the child-related purposes: screening for special educational needs. The sub sections below will consider firstly technical aspects such as reliability and validity, then baseline assessment in screening for educational needs and finally, user evaluation of baseline assessment so far.
3.1 Technical issues and baseline assessment

With regard to the technical qualities of baseline assessment there appears to have been little initial attention from the DfEE although the recent consultation document requests comment on technical aspects (Qualifications and Curriculum Authority, 2000). In 1998, Lindsay and Desforges pointed out that the Schools Curriculum and Assessment Authority recommendations do not refer to either validity or reliability (SCAA 1997c). Lindsay and Desforges emphasised the need for evaluation of schemes used:

*What is common, and depressingly so, is the continuing lack of interest in, or recognition of, the technical quality of the instruments used*

Lindsay and Desforges (1998), 143

Assessment needs to be reliable, or consistent, and to be valid (Roth, 1990). To be reliable a retest should give a similar score for the same child and two different testers should give the same score for the same child, assessed in the same period. Lindsay and Desforges (1998) asked for evaluation of schemes using sizeable samples of 500 - 1,000 pupils, and evidence on reliability over time and across assessor - when two different teachers assess the same child. Reliability is difficult to establish because of the number of baseline assessment schemes in operation.

Validity is a more complex concept and until recently it was argued that there are various types of validity (see Roth, 1990, Miller 1984, Gross 1992). However, it has been argued recently that validity is a unitary construct; the central question is whether an assessment measures what it is supposed to measure – the theoretical construct (for debate about this see Airasian et al, 1999, reviewing Wragg’s *Assessment and Learning*, 1997). For screening purposes predictive validity is
important but however good it is, there are likely to be some false positive identifications, i.e. children wrongly identified, and some false negative identifications, children not identified, (see final section for further discussion).

Returning to the issues of different types of validity, face validity requires that assessment measures what it appears to measure, assessment with predictive validity will correlate with future performance; concurrent validity means the results will agree with other assessments, and construct validity indicates that each item measures the construct it is supposed to measure. The underlying construct of baseline assessment is the range of different skills desirable for a child of five. It is unlikely that all would agree on the desirable skills – see Drummond, (1993) but baseline assessment focuses on those skills which will be needed by the child for reading, writing, maths, language and co-operative learning. Drummond argued cogently that to limit baseline assessment to these skills omits such constructs as creativity (Drummond, 1993).

It may be helpful to consider how predictive validity might apply to the maths items in baseline assessment. Tymms has considered how predictive the maths items were in the PIPs scheme, Performance Indicators for Preschool assessment (Tymms, 1999). Results of this assessment done as children started school were measured against the children's key stage one Standard Attainment Tasks (SATs) at age 7. The items in the PIPs scheme were:

- vocabulary/ concepts such as bigger, most, tallest
- counting a small number of objects (object number concept)
- sums using concrete objects: e.g. if you have three balls and I take one away how many will you have?
• number (digit) identification

and the items in the current baseline scheme (SCAA, 1997, appendix1) map onto these closely apart from an extra sorting item. In Tymms’ study it is unclear what the correlations actually were between performance on these items at nearly five years of age and performance on SATs at seven; he quoted correlations of up to 0.6. with SATs but slightly higher correlations of PIPS baseline against later PIPS reading and maths assessments at age 7 (Tymms 1999, 2). Such correlations imply that the predictive validity of baseline assessment might be expected to be similar insofar as the items in most schemes resemble those of the PIPS scheme.

There appears to be more research on what has predicted reading performance. A longitudinal study showing what predicted future reading performance was carried out in 1987 (Blatchford et al 1987). Assessment of the same children at 4 and then at 7, showed that the following predicted reading at 7 years old:

• vocabulary,
• writing own name or copying
• distinguishing writing from pictures
• letter identification
• reading
• phonological awareness – first sound in word/ rhyme

The most significant predictors appeared to be letter identification and handwriting skills (Blatchford et al, 1987). Tymms (1999) published correlations between baseline skills and attainment at year 2 in reading and maths, and items accounting for the most variance were letter and digit identification. However, the difficulty with all correlational studies is that a third factor may account for both items. For example,
Grogan (1995) has proposed that the factor underlying the correlation between letter naming at four and reading ability at seven may actually be intelligence; she found that knowledge of grapheme phoneme correspondence (that is, the sounds made by letters) and short term memory, shown by auditory and visual sequential memory scores at four years of age, appeared to predict reading ability at seven. Letter naming correlated with intelligence and when intelligence effects were removed, letter naming was not a predictor.

‘Once the effects of intelligence were taken into account, letter naming did not explain a significant amount of variance in reading scores’ (Grogan, 1995, 30)

However, it appears that Grogan used the Draw a Person Test as the measure of intelligence; it could be argued that this is not a comprehensive measure of intelligence (see Richardson, 1991). Additionally, it is possible that both letter naming and the Draw a Person test rely on visual skills and memory. The above indicates one of the dangers of predicting from correlational studies; a third factor may be responsible for variation in both other sets of scores (for a fuller explanation see Miller, 1984). However, bearing in mind the difficulties of prediction from correlational studies, others have found links between preschool phonological awareness and later reading (Stuart, 1995, Lazo and Pumfrey 1996).

Predictive validity of baseline assessment may be measured against Standard Attainment Tests (SATs) and there is a considerable thrust from Ofsted towards schools doing that (Hedger and Jesson, 2000). For screening purposes predictive validity is important; Tymms, writing of the Performance Indicators in Primary Schools (PIPS, Tymms, 1999) argued that that to be useful for measuring value added by the school they should be as predictive as possible – have good predictive
validity. Tymms raises the lack of consistency between schools in the Key Stage 1 data:

'an incidental finding of this research was that - - - there seemed to be some inconsistency from school to school' Tymms. 1999, p24

and in discussion of predictive validity it is worth remembering that the Key Stage 1 data against which baseline is measured may not be reliable and valid.

Tymms has also provided useful information relating to performance and age, showing a steady rise in scores with month of birth. From 48 months of age to 60 months (4 years to 5 years) the change is noticeable. With each month of additional age mean performance on each baseline item increases (Tymms, 1999 and National Sample Data, QCA, 2000). Additionally Tymms notes that month of birth is even more important in contributing to variance in scores (Tymms 1999). Unfortunately this is difficult to draw from the national sample of 1999-2000 (QCA, 2000) but month of birth is confirmed as an important factor in baseline scores by Hedger and Jesson, 2000. They point out that for those born in the Autumn the percentage of baseline scores categorised as low is 22% compared to 47% for those born in summer. Hedger and Jesson group the scores into low which defines pupils scoring below 35% of the total marks and high as those children scoring in the top one third. Unfortunately it is not clear whether these proportions refer to the sample or to the marks – an important distinction in view of variation from school to school and area to area. In any case, the implications are that both age of children at baseline and month of birth are necessary variables to control if any cross comparison between samples is to be made.
At this point it is worth raising a methodological issue about early analysis of baseline; different baseline schemes have been used (Lindsay et al, 2000) and analyses of baseline results use a variety of approaches to the data. For example, it is difficult to compare what Tymms (1999) says about PIPs and baseline assessment with the recent user survey of the various forms used in the national sample (Lindsay et al, 2000; the QCA, 2000). Groupings made by Hedger and Jesson (2000) are low, average and high and such groupings are not comparable with the mean scores which tend to be used in the reporting of the national sample. Not only are different schemes being reported but different statistical methods have been used to analyse the data.

3.2 **Validity for children of different cultures**

This validity, sometimes referred to as cultural validity or ethnic validity (Meisels and Atkins-Burnett, 2000) requires us to ask whether the construct being measured is valued in different cultures. One such construct is that of independence, where a measurement may be that the child puts on his own coat, or uses the toilet/washes hands without help, more highly valued in a western culture. Such issues are rarely considered when assessing young children (Meisels and Atkins-Burnett, 2000).

Additionally, many of the items in the baseline assessment rely on the understanding and use of English (Appendix 1). The Infant Index (Desforges and Lindsay, 1995) showed that children score lower when their mother tongue is not English and the effect varied according to first language. However, when Tymms looked at the correlation of data at seven years of age with the data from the PIPs baseline carried out at four, he found similar patterns of variance (Tymms, 1999). In other words the pattern of skills predicting reading ability was the same.
No data was given for those children for whom English is not their first language in the national sample (QCA, 2000). This is despite the observation by the Office for Standards in Education (Ofsted) in 1999 that

'\textit{a longstanding obstacle to progress is the reluctance of schools and LEAs to monitor performance by ethnic group.}' Ofsted, 1999, para 216

Two schools in the Education Action Zone have 15% of children for whom English is an additional language and other schools in the EAZ have between 3% and 10%. There was no accurate data about the usage or familiarity of children with English, i.e. whether it was not spoken at home, or was to some extent spoken at home.

3.3 \textbf{Baseline assessment in screening for children with special educational difficulties.}

Before considering the current position of baseline assessment and screening for educational needs it is useful to consider some of the debate surrounding early assessment in 1993, when Drummond argued forcibly against it, viewing baseline assessment as 'dead end assessment' (Drummond 1993). She did not believe that baseline assessment showed 'any respect for children's minds'. She also believed that the assessment failed to tap the more complex processes. Her arguments indicate that she viewed assessment as summative rather than formative, and as failing to take account of creativity. Formative assessment is used to inform planning, while summative assessment may be seen as an end in itself or to be used for decision making. Specifically, Drummond proposed that six propositions should be fulfilled before a teacher embarks on baseline assessment. Addressing teachers, she argued that they:
1. (be) 'clear in your mind about the purposes and outcomes: purposes are benevolent while outcomes may be early labelling'

2. 'have clearly in view some kind of picture of what it is you want for your children'
   She cites two baseline schemes as showing children as 'gapingly empty vessels ready for the benevolent input of the reception class teacher'

3. (be) 'clear whether you are assessing for divergence or convergence:' A worthwhile baseline assessment, she argued, would 'encourage educators to reflect on each child's unique characteristics'

4. 'have acknowledged the inescapable truth that there is an emotional dimension to assessment, both for the assessor and the assessed'

5. (be) 'clear about the difference between learning and attainment'

6. 'have done some work on the principles which underlie your practice'
   (Drummond 1999, 37)

A review by McCallum in 1999 was positive about Drummond's caveats (McCallum 1999) arguing for the value of taking a stand. However, both generally and on specific points it is hard to justify Drummond's arguments. In general terms assessment is only 'dead end' if it does not lead to planning and this assignment argues that baseline assessment be used to plan for all the children in the class. It is argued also that the specific points she makes are unfounded:

- The first proposition warned of early labelling: a teacher notes that a child does not score at all for speaking and listening and discusses strategies to encourage
her to listen and speak. Is this early labelling or using assessment to inform teaching?

• Her second and third points are related. She warned about seeing children as ‘gapingly empty vessels’. and asked whether the assessment is for convergence or divergence. Assessment for convergence applies to all normative assessment; baseline assessment pertains to the areas prescribed by the National Curriculum, so that this argument applies to the whole of the National Curriculum and all assessment based on it.

• Her fourth point referred to the emotional dimensions of the assessment. Studies of temperament differences in very young children (Kagan, 1988) show that some children are likely to be more easily made anxious than others; however the baseline items are those items occurring in everyday teaching. It is not clear why the child should feel an emotional involvement unless they find the one to one work anxiety provoking. This would be a matter of how individual teachers manage the assessment rather than a feature of the assessment. At best the teacher may be interested in the child's strengths and difficulties, look forward to using the assessment in planning and enjoy the chance to spend some time with each child. At worst the teacher may be irritated by an extra demand on time. However, there may well be an emotional element in communications with parents (see section 5).

• Fifthly, Drummond exhorted the teacher to be clear about the difference between learning and attainment. In answer to that it could be argued that formative assessment enables teachers to learn how to plan for each child’s learning.
Finally, Drummond felt that the teacher should have done some work on the 'principles which underlie (your) practice.' Once again, this applies to all assessment, not just to baseline assessment.

In summary, Drummond's arguments may be applied to all summative assessments of children. This assignment argues that the use of baseline assessment for planning and for early interventions is a formative use of assessment:

'when (assessment) is decisive in some way the assessment is being used summatively and when it is used to adjust or define it is being used formatively'

(Tymms, 1999 p16)

In fact, Tymms pointed out that few assessments are completely either formative or summative (Tymms 1999) and it must be acknowledged that if baseline assessment is used as an end in itself, and not to inform planning, then Drummond's arguments would have substance.

Wolfendale in 1993 (OMEP UK, 1993) raised points which applied more specifically to baseline assessment; they will be outlined briefly here and the final section of the assignment will return to them in the light of the project with the Education Action Zone. She outlined the purposes of baseline assessment which are to establish a profile of attainment of all children, to attempt to measure and estimate preschool experience, to identify children who may have special learning needs, and to use the information as a basis for measuring future progress (Wolfendale, 1993, 13). She laid down the principle that:

'first stage assessment should be in children's best interests' Wolfendale 1993, 11
and pointed out that the information should be used to plan intervention and to target resources – used formatively, in other words. A list of these questions about baseline assessment are in appendix 3. She asks about its accuracy in the identification of special needs, its relationship to preschool information, and its content. Questions 2, 3, 5 and 10 relate to the formative use of assessment – for planning.

Despite the issues surrounding validity and reliability, user feedback from schools and LEAs appeared to find the current scheme useful in identification of special educational needs (Lindsay et al, 2000). In 1998 Lindsay and Desforges said of baseline that it should:

'accurately identify children who require additional assessment and extra attention - - children having difficulties, whatever the cause, and link with the school's SEN register and scheme for identification and assessment of children under the Code of Practice.' Lindsay and Desforges, 1998, 140

They asked that the scheme should provide a profile of the child's strengths and difficulties and there should be evidence of its validity. Teachers should be able to:

'make initial judgements' and it should be 'linked to planning for intervention': ibid, 141

Baseline assessment is the first step in 'assessment-intervention process' (p141) but other information should be used. Lindsay and Desforges have drawn up a model of the use of baseline assessment for special educational need (Appendix 2) and have summarised criteria for using baseline assessment as a screening tool (Lindsay and Desforges 1998). These were:
• importance: language development, maths, behaviour, concentration were all considered important. (This appeared to be in contradiction to Drummond's view of what was important – creative intelligence)

• Simplicity and acceptability: it should be easy to carry out and acceptable to the people who are using baseline assessment (their recent evaluation indicate a fairly high level of acceptance and approval by users – Lindsay et al 2000)

• Accuracy, sensitivity and specificity: it should pick up those who have difficulties accurately and not yield an unacceptably high number of false negatives and positives. At this point it is useful to expand briefly on the idea of false positives and false negatives referred to by Lindsay and Desforges(1998). In most group screening programmes some will be picked out who should not be and some will be missed who should have been identified. A useful analogy is that of screening in early pregnancy, where false positive identification or false negative identification are translated into probabilities. A false positive in this case would be a child identified by the baseline score as having difficulties, or not having 'school readiness', see discussion (Brookes Gunn et al, 2000). A false negative would be a child whose score did not identify difficulties but who was identified later as a child with special educational needs. The discussion will consider these kinds of screening errors.

• Reliability: assessment should give the same results with different operators and on a test retest basis.
• Further assessment and intervention should follow and the intervention should be focused on the area of difficulty. The authors made the point that a profile is more useful than an overall score (see later discussion in method section). The assessment will then be a true formative assessment and form part of the teaching.

• Lack of ambiguity: it should be clear what is being assessed. This referred to the underlying constructs and they have summarised some of the research on stage theory of development; although there are correlations between assessment and later development these are for groups of children, not predictions for individuals; many other variables, they pointed out, affect children's rate of development. There are discontinuities in development as well as continuity. They have drawn on a model incorporating the interaction between child and environment (Wedell and Lindsay 1980).

Lindsay and Desforges (1998) warned that the two purposes of baseline assessment – to identify children's needs and to use as a measure for later achievement for the school to claim value added - might not sit easily together. They have used the term 'inherent conflict' (Lindsay and Desforges, 1998, 17) and in the evaluation of accredited baseline assessment schemes (Lindsay et al, 2000) teachers indicated a 'tension' between the two functions. For teaching and learning purposes the assessment is probably best done over time and linked to teaching; its use lies in determining where the child is and where s/he may reach – what is the zone of proximal development. For value added purposes an early score is most useful but is not designed to help planning for learning, as it may underestimate the child's capabilities.
Some of the issues raised in this section have related to the ethos of early assessment, and some to the more technical aspects of how it is designed, and validated. How such issues relate to baseline assessment and consultation in the education action zone are discussed in the next section.
4.0 The use of baseline in screening and the criteria of Wolfendale and of Lindsay and Desforges

In relation to this project, Wolfendale's concerns voiced in her 1993 OMEP paper appeared to be unfounded. It did not appear to lead to early labelling; in fact, reception teachers and special needs co-ordinators were quick to remark on progress and in all of the schools some children were removed from the early intervention list. Of the pupils identified in the Autumn term (October, 2000) between one and four in each school has made sufficient progress not to warrant further intervention or consultation. Others remained on, or were moved formally, to a stage on the Register for Children with Special Educational Needs. Whether teachers teach to baseline assessment is a more complex question and given that there is debate about whether children should redo baseline assessment, there may be teaching with those items in mind.

Wolfendale also raised reliability and validity; in nearly half the screenings there was one child not identified by baseline assessment who was raised by the teacher the following term. Such children may be classified as false negatives, that is, not identified for consultation and early strategies. This may be less likely to happen if baseline assessment is used as a basis for calculating value added because there is an inherent interest in keeping the initial score as low as possible. It is difficult to estimate the converse, children wrongly identified as having special educational needs by baseline assessment; it could be argued that the strategies put in place were successful in enabling the child to make progress and move off the consultation list. To identify firmly false positives it would be necessary to implement strategies for an intervention group of children identified by baseline assessment and compare
their progress with a matched control group without intervention – an ethically questionable manipulation. It is argued here that any screening programme is likely to have failures in identification and that reviewing the class the following term should reveal such false positives and negatives.

A further issue is how to measure reliability in view of the number of schemes approved. It is easier to establish the reliability and validity of one scheme. It is worth looking at the recommendations from Wilkinson and Napuk when discussing baseline assessment schemes for Scotland:

'A single national scheme is preferable to an accreditation process - -(which)- - promote an inefficient use of resources by encouraging each authority to devise local schemes' (Wilkinson and Napuk, 1997, 28)

MacKay, however, discussing the baseline assessment schemes which have been piloted in Scotland indicates that one baseline scheme may be 'over prescriptive' for Scotland which has not a prescribed national curriculum (MacKay,1999, p87). In England and Wales, for national consistency and to measure value added, Lindsay and his co authors concluded that one scheme would be preferable (Lindsay et al, 2000). Current plans are to delay assessment until the end of reception year and it may be then be one national scheme in keeping with the foundation stage goals (QCA,2000).

Another of Wolfendale's questions is how far baseline assessment replicates existing knowledge of pupils. In this project the use of information from preschool sources varied in the EAZ. Where the child came from the nursery on site informal discussion usually took place. In two of the seven schools there was supply cover for a meeting between reception teacher and nursery teacher. However, pupil tracking records
were not with the reception teacher when baseline assessment was carried out in two of the first schools. In one school extra educational psychologist time was put into the school to ensure hand over systems were set up. Where staff had changed or the child came from another preschool provision handover of information was variable. In four of the seven schools there was staff turnover in the reception year. During consultation sometimes the reception teacher was able to produce the nursery records, but had not always been able to read them.

To summarise, systems of pupil information transfer were variable and depended on continuity of staff and staff relationships. It is argued here that uniform systems of pupil tracking record transfer should be in place over the EAZ and that transfer meetings should be formally timetabled for the future. In the absence of systematic use of preschool information the baseline screening served as a system to look closely at the skills of each child, to discuss in some depth the children causing concern, to put in place strategies which would help, to share concerns with parents, and to monitor progress and review the following term.

One of Wolfendale's questions to remain unanswered is whether the content items can be sufficiently comprehensive to enable a full and fair profile of the child's development and progress to date to be constructed. In answer to that the new foundation goals appear more comprehensive than baseline assessment and, as assessment is to be at the end of the reception year, it will be based on more thorough knowledge of the child. A second question she asked was whether we know enough about young children's learning to have faith in Baseline Assessment to be accurate. In answer to this we have feedback such as that collected by Lindsay et al (Lindsay et al 2000). The assessment of the new foundation goals may be more
valid and reliable but until it is published and has been validated this is an unanswered question.

Lindsay and Desforges (1998) suggested the following criteria for this first assessment: importance, simplicity, acceptability, accuracy, sensitivity and specificity. In this small scale project there was variation in how important baseline assessment was felt to be. In three of the EAZ schools the reception teachers raised the baseline scores with special needs co-ordinators before this consultation exercise. Fewer than 25% of psychologists in this service (survey at County Service Meeting, September 2000) asked for baseline scores when discussing children of five and six and this is supported by the findings of Lindsay et al (2000) of limited EP use of baseline. Of the LEAS responding 50% said psychologists used it. In the EAZ the profile of baseline assessment and of early screening and intervention has been raised.

Simplicity appeared to be a feature of the current baseline assessment and the recent user evaluation carried out by Lindsay et al, 2000, indicated its general acceptability. However, with regard to accuracy, specificity, sensitivity and reliability, these are areas where the present scheme may be lacking (Lindsay et al 2000). These are all facets of technical excellence and this is one focus of the present consultation document (DfEE 2001).

Initial feedback from school staff indicated that they saw baseline consultation as useful. Special Needs Co-ordinators saw baseline assessment consultation as a valuable way of getting to know the new children early and reception teachers wished to maintain the time dedicated to consultation about new pupils. It was a simple system and, once time was formally allocated to it, was an automatic reminder of the need to discuss special needs in children new to the school. However, two
reservations should be borne in mind when considering teacher responses. The first is that the time for consultation was not usually taken from time allocation to the school. It is possible they would have been less enthusiastic had they felt educational psychology time was spent on this instead of what may have been perceived as higher level needs of older children. The second caveat is that initial feedback from staff was in the form of response to questioning from the educational psychologist. Such qualitative evaluation is subject to experimenter bias, when responses are likely to be influenced factors such as the teacher knowing the questioner, the need to maintain relationships, and other demand characteristics of the situation (Rosenthal 1985). Hammersley (1996) argues that such data is still useful despite this but points out that it:

'makes the setting unrepresentative'
Hammersley 1996 p163

so that ecological validity is compromised. In other words, it would be wise to be cautious about generalising from the responses of these teachers: greater validity would arise from an independent and anonymous evaluation.

4.1 Parental involvement

Of the disadvantages the one which is outstanding is the possible impact upon parents. It is interesting to see how the recent DfEE publication for parents explains baseline assessment:

'it's not a formal test -- it's not something you or your child should be worried about'
DfEE, QCA, (2000) p15
At the start of the consultations some teachers did not appear to be routinely reporting the results of baseline to parents in several of the schools and one experienced teacher said she did not know this was obligatory. Reporting a child's difficulties in certain areas may raise anxiety in parents but some very positive reactions were reported by teachers. One such was the mother of TG whose scores were zero for Reading C, the phonological and sound awareness items and one for letter awareness, Reading B, despite doing relatively well in the other areas. Strategies discussed for TG were extra teaching in rhyme, and onset. When the reception teacher reported the consultation and recommended strategies to TG's mother she said:

'I'm glad you've picked up T so quickly'. We know the EP. She wrote the report for my older daughter's statement and she still can't read at 14. T is exactly the same.'

TG had a different surname so had not been linked with her older sister in consultation by either psychologist or class teacher. Following this feedback from the class teacher and psychologist discussed the use of rhyming and phonological type work to use with TG and the psychologist allocated time to spend with TG on early assessment the following term.

One important aspect of communication with parents is explanation of misleading scores yielding a false positive or false negative. False negatives may occur because a teacher is unsure and gives the child the benefit of the doubt. It will be important to explain the possibility of the score being misleading and for newly qualified teachers in particular it may be useful if the educational psychologist and special needs coordinator can model or role play an parent interview with the teacher.
Teachers were particularly concerned about children with high levels of emotional and behavioural difficulties. Parental involvement was particularly important in the case of children with emotional and behavioural difficulties but it was felt by teachers and parents that little was available for such children; one recommendation would be that further research should investigate what can be done to ensure successful transition to school for such children. Finally, although reception teachers reported positive perceptions of parental reactions, it should be noted that parents have not been directly questioned, and that a number of parents had not attended the meeting with the class teacher. Baseline assessment interviews did appear to increase the teacher-child, teacher-parent dialogue which Wolfendale pointed should be one end of baseline assessment (Wolfendale, 1993). However, as with teachers' feedback, more accurate feedback would result from an independent and anonymous survey of parents after the baseline parent-teacher consultations. The author would argue that for strategies following early consultation full parental involvement is necessary. The vast majority of early intervention programmes rely or even manipulate parenting (Brooks-Gunn et al, 2000) particularly if the targets are language related or to do with behaviour. The efforts to involve parents in the United States in early intervention programmes focused on literacy are described by Brooks-Gunn et al, 2000). The authors consider the idea of 'school readiness', which

'can be interpreted as preparing families and parents as well as children for school'

Brookes-Gunn et al, 550

and they point out that the great majority of early interventions focus in some part on parent-child interaction. The U.S Department of Education’s Goals 2000 programme had both school readiness and supporting parents as the first educators of their children. ‘Social investment’, i.e. additional resourcing such as that of the education action zone, combines with, and should support parental involvement.
4.2 Baseline data in the Education Action Zone

Unplanned consideration of the data raised certain questions about procedures and training for assessment. Although inspection of means indicated that the mean of total scores for the EAZ schools was below the national mean for children of 5.0 to 5.3 there were considerable individual variations within the EAZ schools. Schools sp and ch showed means appearing to be better than the national sample and in the case of schools A enquiry showed that baseline assessment had been carried out by a learning support assistant and it was possible that her assessment had been more positively skewed. Moderation is yearly but informal discussion indicated variation in scoring procedure particularly in items testing more than one skill. It was unclear if all staff had received training and two new teachers were unaware that the Personal and Social Development items are differently scored from the other items, for example. Furthermore there was variation in time of assessment. Some schools carried it out on week two after intake and some on week seven (see Hedger and Jesson, 2000). Additionally, technical faults in the test construction may contribute to variation, such as items which assess two constructs. Some teachers may give credit when only one of these skills is achieved, particularly if they feel the items depend on the child's ability to use and understand a certain type of language.

With reference to the data yielded it has been difficult to interrogate both national and local data (see section 2.3.3). For example the variability in EAZ schools is noticeable but cannot be compared to national variance as standard deviations are not provided. It is difficult to know what effect the scores of children with English as an additional language may have. There is a clear government recommendation that data is collected for children who have English as an additional language to raise
achievement of ethnic minority (Ofsted, 1999) but it was not possible to gain this information, despite a sizeable local recent refugee population and an established Asian and Afro-Caribbean population.
5.0 Conclusions and recommendations

Firstly, it would be useful to have formal evaluation of consultation following baseline assessment, by psychologists who did not implement the procedure using surveys of both teachers and parents. It would also be useful to follow up children who were the subject of consultation and compare them to a control group from schools similar to those in the action zone. For such consultation using assessment data, the data needs to be valid and reliable. (At this point it is worth noting that the more effective early intervention is, the less predictive baseline assessment will be of long lasting special educational needs.)

The new assessment at the end of the Foundation Stage should be carefully trialled and could be evaluated against other assessments such as Key Stage One and Two SATs, and against other assessments which have already been validated, for example assessments from the National Foundation of Educational Research (NFER), or Wechsler (Wechsler Objective Numeracy Dimension, Wechsler Objective Language Dimension, and reading assessments already validated. It would be very much easier to look at local and national data if the same conventions were followed in the gathering, grouping and analysis of the data.

Secondly, for many parents this will be their first focused interaction with their child's teacher. Baseline assessment has presented the teacher with some data about the child and she wants to do two things; checkout any hypotheses she has formed about the child's abilities or difficulties, and involve the parents in the strategies she wants to put in place. Additionally, foolproof systems should be set up to ensure that records are transferred to reception teachers and that time is set aside for discussion of the children transferring to school so that parents are aware that records have
transferred and been read. Parental feedback on interviews with teachers after baseline assessment would be helpful.

Thirdly, in this small scale study, teachers and head teachers remarked repeatedly that they did not have interventions or strategies for children with high levels of emotional and behavioural difficulties. The statutory procedures were lengthy and there was exclusion in some cases. For pupils and staff in the EAZ it is recommended that research takes place on how best to ensure transition for children with marked behavioural difficulties to school from nursery and to ensure appropriate and early involvement of other agencies, such as behaviour outreach teams, and education officers.

Fourthly, it may be useful to involve educational psychologists in the design and evaluation of the assessment to take place at the end of the Foundation Stage. Their training and use of assessment makes them particularly fitted to take part in this. However, of those LEAs which responded to the survey by Lindsay et al (Lindsay et al 2000), only about 50% said educational psychologists were using baseline. What might be the reasons for so few educational psychologists using it? By training psychologists may be wary of assessments. In 1994 Bagnato and Neisworth found that only 4% of the psychologists they surveyed supported the use of norm referenced standardised intelligence tests with young children. Psychologists may be wary because of false positives and negatives, or of parental anxieties. Educational psychologists may traditionally be less involved at these early stages; they may have doubts about the construction of the assessment using sound psychological principles. High level needs may use up available psychologist time and absorb the attention of reception teacher, special needs coordinator and head teacher. If children with severe emotional and behavioural needs are part of a new intake it is likely that attention to less severe needs will be limited.
However, it is argued early interventions require the use of screening procedures and the use of an existing compulsory assessment is practical and economic. At present this is baseline assessment, but assessment of Early Learning Goals is to be introduced (QCA, 2000). Assessment for intervention is what Meisels and Atkins-Burnett (2000) refer to as the fusion of assessment and learning; assessment is a first step in a dynamic process, calling on information from numerous sources, and it is closely tied to intervention. In the light of the above project it appears useful to involve psychologists in consultation following such assessment, using a similar procedure to the one outlined above.
References


Grogan, S., (1999) Which cognitive abilities at age four are the best predictors of reading at ability at age seven J.Res. Reading (1995) 18(1) 24-31


43

Hedger K, Jesson D, (2000) The Numbers Game : measuring pupil progress from Baseline to Key Stage 1 Publications and Design Unit, Shrewsbury


Lazo M.G. Pumfrey P.D., (1996) Early predictors of later attainment in reading and spelling Reading 30(3) 5-11


Oxfordshire Early Years Development File Oxfordshire County Council Education Service, Early Years Team

Office for Standards in Education, 1999, Raising the attainment of minority ethnic 'pupils. London: Ofsted Publications Centre


Richardson, J.T., ed (1996) Handbook of qualitative research methods Leicester: The British Psychological Society


School Curriculum and Assessment Authority (1997a) *Baseline Assessment Scales* CM/97/809. London:SCAA


Appendix 1 Baseline assessment scales

Individual Record Sheet

Reading A: Reading for meaning and enjoyment
1. Holds book appropriately while turning the pages and retelling the story from memory (2)
2. Able to predict words and phrases
3. Uses memory of familiar text to match some spoken and written words
4. Reads simple texts

Reading B: Letter knowledge
1. Recognises his or her own name
2. Recognises five letters by shape and sound
3. Recognises 15 letters by shape and sound
4. Recognises all letter shapes by names and sound

Reading C: Phonological awareness
1. Recites familiar rhymes
2. Recognises initial sounds
3. Associates sounds with patterns in rhyme
4. Demonstrates knowledge of sound sequences in words

Writing:
1. Distinguishes between print and pictures in own work
2. Writes letter shapes
3. Independently writes own name spelt correctly
4. Writes words

Speaking and Listening:
1. Recounts events or experiences
2. Asks questions to find out information and listens to the answer (2)
3. Makes up own story and tells it
4. Makes up a story with detail and tells it to a small group, and listens to stories (3)

Mathematics A: Number
1. Sorts sets of objects by given criterion and explains sorting (2)
2. Counts objects accurately
3. Shows awareness of using addition
4. Solves numerical problems using addition and subtraction

Mathematics B: Using mathematical language
1. Can describe size
2. Can describe position
3. Recognises numbers to 10 and writes 1-10 (2)
4. Can explain an addition sum

Personal and Social Development:
1. Plays collaboratively
2. Is independent and keen to contribute (2)
3. Concentrates without supervision for 10 minutes
4. Expresses own opinions with a range of adults

Where a number in brackets follows an item it assesses more than one skill

School Curriculum and Assessment Authority (1997a) Baseline Assessment Scales CM/97/809. London: SCAA
Appendix 2: A model of baseline assessment

Lindsay G, Desforges M. (1998) Baseline Assessment: Practice, Problems and Possibilities: London, David Fulton Publishers (Figure 3.3)

Figure 3.3 An elaborated model of a baseline assessment as a means of the early identification of special educational needs
Appendix 3 Wolfendale’s questions about baseline assessment in her report for OMEP 1993

1. Can it lead to early labelling of children?
2. Will teachers teach according to the way in which children respond on Baseline Assessment measures?
3. Can Baseline Assessment information accurately identify special educational needs on entry to school?
4. Is it fair to set children tasks to do on which their performance will be judged?
5. Is Baseline Assessment information comprehensive and reliable enough a basis for allocating staff and material resources?
6. Is Baseline Assessment the most effective way to gauge the range and type of preschool experience?
7. What is the distinction between an early years record and an profile?
8. Can the content items be sufficiently comprehensive to enable a full and fair profile of the child’s development and progress to date to be constructed?
9. Do we know enough about young children’s learning to have faith in Baseline Assessment to be accurate?
10. Does Baseline Assessment facilitate a teacher-child, teacher-parent dialogue?
11. Is Baseline Assessment a logical move downwards from testing at age 7?

Professional Practice Assignment

Submitted in part fulfillment of the requirements for the Continuing Professional Development Doctorate in Educational Psychology (DEdPsy)

Name of Course Member: Deborah King

Title of Assignment:

Measuring the outcomes of consultation: part of a best value approach.

Assignment number 2

Core Curriculum area to which this assignment topic relates:

The Profession of Educational Psychology and its Context
## Contents

<table>
<thead>
<tr>
<th>1.0</th>
<th>Introduction</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Evaluation of educational psychology services</td>
<td>51</td>
</tr>
<tr>
<td>1.2</td>
<td>Evaluation of service delivery through consultation</td>
<td>52</td>
</tr>
<tr>
<td>2.0</td>
<td>Psychological processes underpinning consultation</td>
<td>55</td>
</tr>
<tr>
<td>2.1</td>
<td>Persuasion</td>
<td>55</td>
</tr>
<tr>
<td>2.2</td>
<td>Attitude and its relation to behaviour</td>
<td>56</td>
</tr>
<tr>
<td>2.3</td>
<td>Helping staff believe in their own efficacy</td>
<td>59</td>
</tr>
<tr>
<td>2.4</td>
<td>The consultation process: process and outcome</td>
<td>63</td>
</tr>
<tr>
<td>2.4.1</td>
<td>Measurement of process :the consultation record</td>
<td>63</td>
</tr>
<tr>
<td>2.4.2</td>
<td>Measurement of outcome</td>
<td>65</td>
</tr>
<tr>
<td>3.0</td>
<td>Developing a service measure of outcome</td>
<td>69</td>
</tr>
<tr>
<td>3.1</td>
<td>A single county Survey in County Y</td>
<td>71</td>
</tr>
<tr>
<td>4.0</td>
<td>Recommendations on measurement of outcome for County Y</td>
<td>79</td>
</tr>
<tr>
<td>5.0</td>
<td>Measurement of consultation outcome as part of service evaluation: future directions</td>
<td>82</td>
</tr>
<tr>
<td>5.1</td>
<td>Future directions for the evaluation of consultation outcome</td>
<td>84</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>86</td>
</tr>
<tr>
<td>Appendix 1</td>
<td>Proposed evaluation of the Educational Psychology Service</td>
<td>88</td>
</tr>
<tr>
<td>Appendix 2</td>
<td>Consultation effectiveness measure</td>
<td>89</td>
</tr>
<tr>
<td>Appendix 3</td>
<td>Measuring change, County S</td>
<td>90</td>
</tr>
<tr>
<td>Appendix 4</td>
<td>School feedback for the Educational Psychology Service</td>
<td>92</td>
</tr>
</tbody>
</table>
1.0 Introduction:

1.1 Evaluation of educational psychology services

This assignment considers how consultative educational psychology services evaluate the consultation they offer. It will discuss the measurement of what happens as a result of consultations and how well the service meets the needs of users. It will be argued that to be useful to the service, measuring the outcomes of consultation should be considered in the context of the overall evaluation measures of the service.

In the second section the assignment focuses on the measurement of consultation outcomes. It discusses some of the research into the psychological processes underpinning change occurring in consultation – persuasion, attitude change and the relationship of attitude to behaviour. It goes on to consider the actual process of consultation; how to analyse consultation and to measure its perceived effectiveness.

The third section considers specific examples of how particular services have approached the evaluation of consultation. Service Y has recently introduced consultation and has been surveyed on how they would measure the outcomes of their consultation; their responses are considered, particularly in the light of how to move towards measuring outcomes of consultation. Recommendations are made on how the service could move forward on this.
The fourth part considers how such measurement could fit into the prevalent management approach to evaluation of psychological services using performance indicators to establish whether a service gives 'best value' to its users.

1.2 Evaluation of service delivery through consultation

In local authorities recently there has been a growth in measurement of services to clients; there are quantitative measurements which look at how much work has been completed – how many reviews were attended, how many statutory assessments completed within time limits. There are also user measurements, that is, how satisfied are users with the service (DETR 1999). Industries which have a readily quantifiable product such as banking and finance, can more easily measure outcomes. In education there has been a public acknowledgement of measurement with published examination results, and formal evaluation by the Office for Standards in Education (OfSTED). Alongside this many educational psychology services have adopted a consultation model of service delivery, and are considering how to evaluate what happens as a result of consultation.

Empirically speaking, it is not possible to isolate consultation as a variable; it cannot be measured in isolation from other variables springing from the school staff, child or family. However, one approach has been to look at the impact of consultation – the effect on the consultee, and the outcome – for the child or student; various studies have done this, particularly in the United States (Erchul, 1987; Hughes et al, 1997; Busse et al, 1999; Farouk, 1999) A review of outcome research outlines the lack of rigour in some studies of effect (Sheridan et al 1996) and calls for measurement using objective indices. The authors also call attention to the link between specifying the model of consultation and measuring its effect; the school
consultation research which they review appears to focus predominantly on behavioural consultation where the teacher and/or parent is the consultee and the consultant the educational psychologist.

Additionally there appears to be what Sheridan refers to as variations of these models in the definition of the process such as the amount of collaboration or directiveness. (see Gutkin 1999 for further discussion of this). Within the United Kingdom Farouk (1999) also makes the point that consultation in schools is varied. However, Conoley and Conoley have a broad definition which would cover most models:

'\textit{a voluntary, non supervisory relationship between professionals from differing fields to aid one on his or her professional functioning}'.

Conoley and Conoley (1982,p2)

What are the processes underlying a successful consultation and how can we measure its effect? Hans Eysenck measured the effects of psychotherapy and found that the rate of recovery was very similar for those having had or not had therapy (Eysenck, 1952). Might there be a similar outcome for consultation and could measurement of effects be approached in a similar way? To look at measuring the effects of consultation it is necessary to look at the actual psychological processes underlying consultation – persuasion, attitude and intention and how these relate to planned behaviour – the macro processes. These are discussed in the next section.

The micro processes of consultation are then considered. The nature of the interactions during consultation have been the subject of analysis through the use
of the Consultation Analysis Record (Bergan and Tombari 1975) and the data can be used to categorise the nature of the exchange of ideas; for example, it may be used to show how directive or collaborative the consultation is (Gutkin 1999).
2.0 Psychological processes underpinning consultation

2.1 Persuasion

Consultation has been seen as a process of influencing others (Short et al., 1991). There is a large body of research on the effect of persuasive messages; specific research on persuasion in consultation (Hughes et al., 1990; O'Keefe and Medway, 1994) is more useful, because generalisation from other areas such as advertising, may not be warranted. For example, sample sizes or participant groups may be too dissimilar. There is also an ongoing relationship between consultant and consultee. Persuasion may be necessary to make the consultation effective. Unless the consultee is persuaded to act, the process of intervention will stop after the consultation (Gutkin and Curtis, 1990)

Persuasion in consultation may be used to:

- limit digression
- increase objectivity
- look for the positive, or the change in behaviour
- move the discussion to the knowledge or skills of the staff
- persuade staff not to exclude
- refer, or not to refer to some other agency

(O'Keefe and Medway 1994)

O'Keefe and Medway further argue that persuasion in consultation is an aid to problem solving and to building relationships, and so should not be seen as manipulative.
They make the caveat that objections which the consultee raises to a proposal could be valid and should be considered and may lead to the original proposal being abandoned, or modified. However, there is an assumption by O'Keefe and Medway that persuasion will be used for a desirable end – and that the psychologist decides on what is desirable; this is in contrast with the emphasis by others on the joint nature of the problem solving and on an open trusting relationship (Gutkin and Curtis, 1990).

If we apply this to an exclusion consultation it is possible that such persuasion, on behalf of inclusion for the psychologists' client, may raise ethical concerns. An area of ethical conflict not explored by O'Keefe and Medway is the reconciliation of the best interests of the client with those of other students and the employment environment of the staff.

### 2.2 Attitude and its relationship to behaviour

Assuming that the consultation reflects jointly held attitudes, there are other intervening variables between attitude and behaviour. The person may not remember their beliefs, or they may remember them but may not see themselves as effective in carrying out planned behaviour. Staff may be aware of their own attitudes and how they relate to their own intention and to their behaviour but may have erroneous beliefs about the attitudes of others – colleagues and pupils. Behaviour may not reflect attitude; research shows that there is a place for reminding people of their attitudes in order to bring about desired behaviour. For
example a head teacher may want a pupil to go to a special school. S/he may believe in inclusion but resist inclusion for this child. Reminder of the attitudes staff may hold can be helpful.

'I know you felt very positive about including x with Down Syndrome'

A similar reminder of attitude was carried out with students by Snyder and Kendzierski (1982) who asked students, known to have positive attitudes to research, to volunteer for quite demanding research experiments; they overheard one of two conversations between confederates:

'I don't know if I should volunteer or if I shouldn't volunteer. What do you think?'

Control group answer: 'Beats me—it's up to you'

Experimental group answer: 'Well, I guess whether you do or whether you don't is really a question of how worthwhile you think experiments are.'

Twenty-five percent volunteered in the control group and sixty-five percent in the experimental group, showing that reminders of attitude are effective.

A further variable intervening between attitude and behaviour is how effective a person feels herself to be. The theory of reasoned action (Ajzen and Fishbein, 1980) argues that the behavioural intention determines (mostly) what a person does and behavioural intention is a function of self efficacy, among other variables. In fact the theory of reasoned action reminds us of how little we can
predict people's behaviour by knowing what their intentions are and what behaviour they plan. However, in a consultation we might work on:

- the person's own attitude
- subjective norm - the person's perception of how others view the behaviour.
- perceived self efficacy

It is likely that perceived self efficacy relates to previous experience in the classroom – a newly qualified teacher with a particularly difficult class may find that she is experiencing learned helplessness (Seligman 1975). In fact there are contradictory findings about the relationship between self efficacy, experience and consultation. Hughes et al found that consultation tended to be sought by teachers with least experience and not sought out by those with most self efficacy (Hughes et al 1990). On the other hand Gutkin and Hickman found indications that teachers who feel in control will choose consultation rather than referral (Gutkin and Hickman, 1988).

It appears likely that this apparent contradiction is because of the juxtaposition of consultation and referral in Gutkin and Hickman's research. A teacher who feels in control but can see a difficulty developing, or a difficulty with other staff, may choose to discuss the situation. A teacher who feels that they will not be able to bring about change - i.e. that they have lost control, and have not the self efficacy to change that, is more likely to wish to refer – to pass the student on. Consultation research focussing on students with emotional and behavioural difficulties, such as that of Busse et al, 1999, may find the choice to consult is made by the teachers who feel that they can maintain the student in a classroom setting. Those feeling that the situation is outside their control are more likely to choose referral. This is in line with Gutkin and Hickman (1988). What such findings illustrate is that self
efficacy is probably a central variable in the choice of consultation, particularly when the student has emotional and behavioural difficulties.

2.3 Helping staff believe in their own efficacy: a case study

There may be a variety of reasons for not choosing or choosing consultation, or for not implementing what was agreed, (Dougherty 1994):

- a lack of perceived control and a feeling of inexperience may act against a teacher choosing consultation,
- experience and high self efficacy may also act against it.
- a further rationale for referral may also be that the teacher does not believe that anything will arise from consultation which she has not already considered, and either tried, or rejected as impractical.
- teachers who believe they have control concerning a specific student problem are more likely to choose consultation services than referral (see above, Gutkin and Hickman 1988).

This points to an initial skewing in the use of consultation. Teachers who do not feel that they are efficacious may not wish to use the educational psychologist for consultation. Referral indicates a belief that the locus of control is to be found elsewhere – that the teacher cannot bring about change. This brings us back to the importance of self efficacy; the following case study illustrates the important part that self efficacy may play in a failing school.
### Case study of a school under special measures

In school improvement issues – where schools are in special measures or of concern to the LEA, the staff are likely to have considerable doubts about their self efficacy.

The staff of a first school in special measures, after initial rejection of the findings and anger, appeared to become quite depressed; the belief of the staff in their self efficacy was low.

An action plan was drawn up by the staff and governors but with a sense of it being imposed from outside. There were complaints from staff and head about the number of visitors and two of the four teachers left. New staff felt ‘constantly visited and under a spotlight’

Staff were reluctant to meet with EP; three meetings were cancelled. The EP focused the first meeting around the school’s Action Plan which criticised early identification of special educational needs. EP and reception teacher worked through baseline screening, jointly identifying children whose scores were low in any areas; the class teacher was encouraged to think of helpful strategies for these children, and the children were reviewed the following term. It was noticeable how much more enthusiastic the reception teacher was about the next consultations on the new intake’s baseline scores. She said that

> *my strategies worked for S—- and P—- but can we talk about D—- moving to stage one of the SEN register?*

It is important to note that the reception teacher felt the strategies were her ideas. She was then able to take credit for the progress of the children and to discuss positively the next term’s intake.

The annual school feedback to the EP service pinpointed consultation:

> *the consultations between X (EP) and staff have been vital — baseline assessment screenings have proved very useful in our early identification strategy. As this is part of our OFSTED Action Plan it has been important*

Staff in a school on special measures are likely to feel that they lack efficacy. It could be argued that the prime aim of consultation would be to recover this sense of self efficacy while implementing the Action Plan. Little useful consultation could take place until the staff started to feel effective. EP consults re evaluation of SEN action plan and criteria with each member of staff, consults with behaviour outreach/advisors and governors. The newly qualified reception teacher takes charge of early screening and intervention.

The registered inspector visits and points out the marking in year 3 and 4 is not up to date. The newly qualified teacher has been doing Sats, and is the Science co-ordinator.

The Head teacher tells the class teacher this delay in marking is not professional. Both the head and the class teacher suffer a drop in self efficacy; consultation follows with the EP; the class teacher needs to recover self esteem and a feeling of self efficacy and the head needs some support in staff development and management. EP liaises with adviser to support.
In addition to the reasons Dougherty outlines (start of this section, Dougherty, 1994) there are other intervening variables for not implementing what was agreed in consultation:

- staff may have forgotten the attitudes they hold or do not see how they relate to the planned action (Snyder and Kendzierski, 1982)
- the attitude doesn't correspond with the desired behaviour (O'Keefe and Medway, 1994)
- the person does not believe they can do it (Fishbein and Ajzen 1980)
- The teacher may not hold the desired attitude, or believe that it contradicts the norms of management and staff.

Effective consultation needs a 'sympathetic and accurate understanding of the consultee's circumstance (O'Keefe and Medway 1990,183). It is important to know why a teacher does not want to consult – possibly because s/he doesn't feel effective or because s/he doesn't believe the consultation will result in any change. Factors working against effective consultation would be the psychologist holding the wrong beliefs about the consultee’s attitude, or trying to be too openly coercive. Two common mistakes in consultation are the assumption that attitude change is needed and the assumption that the basis of attitude is known (O'Keefe and Medway 1997).

Domination appears to be present in many consultations:

*Consultation often looks like 'domination rather than collaboration'*

(Henning-Stout 1993, p37.)

The issue of direction and collaboration in consultation is addressed by both Erchul (Erchul1992) and Gutkin (Gutkin,1999). Both conclude that there is no evidence
from analyses of consultation of directiveness at the expense of collaboration.

Gutkin discusses a model of consultation on an orthogonal axis.

(Gutkin, 1999, p181)

Such a model gives rise to four approaches: collaborative-directive, collaborative-non-directive, coercive-directive, coercive-non-directive.

Gutkin suggests the collaborative-directive model may be the most helpful approach but he also suggests that to be effective, perhaps consultants will move between approaches.

To sum up, O'Keefe and Medway point out that to be effective, consultation

- should be collaborative
- may be directive
- should be specific; clear specification of behaviour plans are seen as useful although they may not bear any relationship to outcome for the child/student
- should be followed up because of regret after decisions.
- may need to boost self efficacy
2.4 The consultation process: process and outcome

2.4.1 Measurement of process: the consultation analysis record

In the United States universities run graduate classes for psychologists in consultation skills and empirical research on consultation and its effectiveness has taken the form of measuring what actually happens – what is said during consultation – and relating this to impact and outcomes. In depth research may be carried out on consultations using the CAR – the Consultation Analysis Record (Bergan and Tombari 1975). This enables coding of the consultation in terms of how many verbalisations are made by the consultant and consultee and what form those verbalisations take – questions, or eliciting information, requesting validation, commands, instructions, requests.

Hughes et al (1997) hypothesised that the greater use of open questioning by consultants would lead to a rating of greater effectiveness and that this would also be enhanced by the greater use of questions about consultees’ thinking – a collaborative approach. Forty one graduate students, trained in the consultation course took part in at least two consultations with teachers, focusing on a specific student as the client. The evaluation took the form of the CEF, the consultant evaluation form, filled in by the teachers, the consultees. This is 12 statements of consultant effectiveness rated on a 7 point scale (developed by Erchul 1987). The interviews were coded using the CAR. Correlations were modest and mostly below significance level; there was a significant correlation between the evaluation and the number of inference and accepted questions; inference questions indicate a collaborative approach and accepted questions are those responded to, as opposed to evaded or refused.

Hughes et al. themselves point out that:
‘these results contribute to a body of literature on verbal processes in consultation that has produced small and inconsistent findings.’ p292

and that most of the findings are null where research tries to relate consultancy to outcomes for the child.

The most noticeable findings relate to consultees who attempt to direct – that is attempt to control the consultation. Such consultees are less positive about the consultation and less likely to carry out the tasks associated with consultation. These findings must be of considerable interest to those psychologists whose schools may have resisted the introduction of consultation, or where the consultation is with a head teacher who takes a directive part, over exclusion, for example.

An interesting study was carried out by Busse et al analysing the content of consultations about children with emotional and behavioural difficulties – often perceived as a difficult consultation. (Busse et al 1999) They relate the coding of consultation sessions to the outcomes for the client – the student. The study used twenty seven consultations in groups of three; the first interview identified the problem, the second identified a treatment plan and the third evaluated it. The authors predicted that CAR would show that:

- consultants demonstrate more control of the consultation and that control accounts for significant variance in treatment outcomes

The findings supported the first prediction:

- consultants did show more control of the consultation process than the consultees but the realtionship between this control and the outcomes was complex.
• The control did not predict more favourable outcomes, nor did the planning of behaviour specifications and strategies.

• There was no relationship between the consultees’ perceptions of the consultants’ effectiveness and the treatment outcomes – the progress of the student

The authors suggest that consultants engaging in excessive planning and specification may have a negative impact. [However, educational psychologists may find that in consulting on how to bring about behavioural change it is likely that that the greater length of time and detail are spent on just those behaviours which are very difficult to change]. The authors spend some time discussing the methodological limitations of the CAR and other similar procedures for analysing verbal behaviour. They point out that:

• it does not measure verbal interaction but discrete utterances,

• the strongest multiple regression left 65% of the outcome variance unexplained,

• they had not controlled for severity of behaviour,

• some measures were teacher ratings.

2.4.2 Measurement of outcome

The study discussed above, (Busse et al 1999), appears to be a particularly important one for services who wish to adopt a consultation approach in all activities. There was no relationship between the consultees' perceptions of the consultants' effectiveness and treatment outcomes for the students; this may either mean that consultation does not affect what happens to the student, or that how it is perceived by the consultee bears no relation to what happens to the student. A similar study controlling for severity of pupil difficulty and scrutinising the other
causes of variance in outcome would be of interest to psychologists using consultation.

Research into this consultation process and outcomes does not show a clear relationship. In common with other measurements in the field of social psychology it may be that psychologists will be reluctant to state clearly what they want to measure, i.e. they would like naïve participants – an ethical problem. Some commonly used measurements of impact and outcome are listed below:

- measuring attitude change: it may be that attitude change has to occur, or behaviour of one or more parties, or the consultee's perception of consultation (see O'Keefe and Medway, 1997). It is rarely stated explicitly that this is an aim of consultation and so it is unusual to see a measurement of consultation with a scale on attitude change. This is probably because it is not made clear that one aim of consultation may be to change staff attitude, although it is often the staff aim for the pupil.
- measuring change in learning or behaviour of the child: it is relatively simple to measure what the teacher perceives as the change with a rating scale or by using an outcomes and review approach (see Dickinson, 2000).
- measuring change in behaviour of staff/school; here it is likely that the scale will use an indirect form of wording – e.g. how have policy and practice changed?
- measuring self efficacy – in either child or teacher: it is likely that the scale will be couched in terms of self confidence or, for the teacher, the certainty of bringing about change (Hughes et al., 1990. Hughes et al., 1997)
- measuring the perception of the consultee about consultation: there should be a question about how happy the consultee is to continue with consultation about other pupils, groups of pupils, or difficulties in his school (see the model
All of these are more difficult to measure empirically than the process measurement discussed above – the mechanics of the interaction between teacher and educational psychologist. However, the process measurement applies to the impact or perceptions of the consultee and do not necessarily relate to the outcome for the client – the student, or child discussed.

In this country Farouk has pursued the perceived effectiveness from the consultant's view. He has questioned educational psychologists about what they perceive as affecting the take up of suggested strategies (Farouk 1999). The most important factor was seen as the teachers' degree of ownership and commitment and the most frequent comment was the need to work collaboratively. The usefulness of EP advice was not mentioned in these answers - but presumably, following the article by Woods (1998) *Okay, then: what do EPs do?*, which questions the usefulness of educational psychologists, psychologists must wonder how their advice is perceived.

It is necessary to query whether we can generalise from Farouk's findings to the whole profession for two reasons:

- the return to Farouk's questionnaire was 50% and perhaps a skewing of the sample occurs in that more positive EPs returned the questionnaire.

- having made a decision to work in a consultative way, a psychologist may reassure herself that the choice is correct (see, Theory of Cognitive Dissonance, Festinger 1957).
There are two salient facts which appear to stand out in research on effective consultation; firstly, various personal attributes contribute to success: it is likely that being non judgemental, genuine, having good communication and active listening skills and empathy all contribute to successful and fluent exchange of ideas and joint problem solving (Dougherty 1994).

Secondly, the fact that a consultation is perceived as effective by the consultee does not necessarily mean that it has had an effect on the outcome for the student. There does not appear to be a relationship between how effective the consultation is perceived as being by the consultee and the outcomes for the pupil (Busse et al 1999).
3.0 **Developing a service measure of consultation outcomes**

Two draft service approaches to measuring outcomes are considered. They distinguish between

- the impact
- the outcome

The impact measures the effect on the consultee and the outcome is what happens as a result for the pupil. Examples of services which have moved towards a formal measurement are B and S. (Appendix 3)

Services such as B are considering asking post consultation in a measure of impact:

*How able do you feel to deal with the situation now?*

A measure of the outcome would be:

*How have things changed for X (pupil)? What has happened for X which is different following consultation?*

S. has developed a series of questions which appear to emphasise impact. Pre and post consultation rating scales are to be filled in by the consultee.

1. *How confident do you feel about being able to manage the situation now?*

---

1 Services are anonymised as many schemes are in draft
2. How confident do you feel that you could use the consultation outcome to manage a similar situation with another pupil or group of pupils.

3. In what ways did the consultation session change your understanding of the problem.

4. In what ways did the consultation change your professional behaviour.

5. What was it that the educational psychologist did or said which was particularly useful.

6. Was there any additional service the educational psychologist could have helpfully provided (if so, please say what you feel should have been).

(These questions are taken from draft proposals to the Educational Psychology Service in S, see Appendix 3)

Scrutinising these questions in the light of what we know about the consultation process, questions 3 and 4 may not contribute to feelings of efficacy, implying possible poor initial understanding and management on the part of the consultee. Indeed, it might be argued that such questions contribute rather to a feeling of learned helplessness (Seligman 1975): the consultee may be left with,

I did not understand, I had poor professional management, but following consultation, both of these improved. What brought about the improvement?

Consultation.

Service B asks:

How able do you feel to cope with the situation now?
Such questions do not really measure the outcome for the student. The study by Busse et al (1999) indicates a lack of measurable relationship between perceptions of the consultée about consultation and outcome for the student. However, they do attempt to measure growth in confidence in dealing with the problem. Will any question which reminds a consultée of initial lack of confidence, increase confidence?

3.1 **A single county survey in County Y**

For this assignment educational psychologists in county Y were questioned about how they would like to measure outcomes of consultation. Their service had recently moved to consultation but implementation has been variable. The Principal Educational Psychologist wishes to develop a measure of the outcomes of consultation, and is also currently exploring a move to using performance indicators of the best value approach.

**Method**

At a county service meeting all psychologists were informed that the author was gathering information on how to measure the outcomes of consultation, and that this had been requested by the Principal Educational Psychologist. They were also informed that this would form part of an assignment for the Continuing Professional Development Doctorate. A form was sent to each psychologist in which s/he was asked to write down their thoughts on how to measure outcome of consultation.

The question was made open in order to place as few constraints as possible on thinking.
I would be very grateful if you could think about how, after consultation(s) we could measure what happens for the client – the outcomes for child, school staff, parent, partnership, LEA

At the following county meeting psychologists were reminded of this and forms given to those who had not returned them.

Of twenty four psychologists, eleven returned forms, one gave a project done in training on evaluation of consultation and one emailed a response - similar to the 50% return for Farouk (1999). Three of those responding phoned to ask what sort of ways to measure outcome might be used. The reply was that they should include whatever they thought a relevant measure of outcome.

Analysis of replies

All replies considered outcome measurement. However the outcome measures fell into two categories; those who considered the evaluation of consultation to relate directly to the client (child, student) and those who considered it to apply to client through strategies recommended to the staff. One reply considered impact on consultee– measuring the perception of the consultee:

'staff feel they have acquired skills they could use with other children'

1 Ensuring strategies recommended were carried out

Six of the replies related to how to ensure that strategies recommended were carried out – e.g. speech therapy requested, behaviour management strategies and one quote is typical of these answers:

'to ensure that the teacher recognises their part'
2 Placing on the Code of Practice for children with Special Educational Needs

Five responses focused on appropriate placing on the Code of Practice register, and four of these said that movement on the register should be measured; moving down a stage would be viewed as a positive outcome for the child as would ceasing to maintain the statement.

3 Interagency work

One response was geared towards the ensuring of interagency work through consultation and as a result of consultation a sheet was filled out relating to action to be taken by the various support services. This psychologist had developed this line in his previous authority. This related to the carrying out of actions agreed in a meeting.

4 Parental follow up

One response noted that this should be done as part of measuring outcome.

5 Measurement of consultation in project work

All replies assumed that the child was the client, not the consultee. One which very much focused on the child as client is the reply from the psychologist responsible for co-ordinating a joint agency project. This is given in some detail because it exemplifies a project which has chosen to adopt client measures of outcome.

Health, Social Services and Education have funded and staffed this project, known as the Tier 4 project. Pupils who are in the looked after system, with a statement for emotional and behavioural difficulties and currently in county specialist provision
were the focus. The aim was to maintain these children (the figure is about 10 and the ages were from 10 to 14) who were presently at risk of requiring out of county independent provision because of the difficulties they were experiencing.

The educational psychologist is the co-ordinator. The project employs an educational psychologist, a social worker, a psychiatrist, an art therapist and drama therapist, as well as staff from the special school for children with emotional and behavioural difficulties. Rationale for funding came from the argument that the children's needs were better served by maintaining them in their local community and that money saved from out of county provision would pay for what is relatively expensive and out of the ordinary provision.

One year into the project the psychologist made the point that desirable outcomes should be specified at the start and these may include all or some of the following:

quantitative, generally available measures

- the Social Development Quotient
- attendance
- daybook entries
- privileges
- incident sheets
- children homes – record of assaults/ abscondings
- mainstream records of behaviour
- national curriculum levels
client specific measures
these are to be discussed at the point of referral. In other words an outcome is
specified at the start
• number of lessons attended
• episodes of self harm
• sessions of therapy attended
• number of phone contact
• number of unsolicited phone contacts
• engagement reflected in comments/ requests for staff to involve themselves in
  other areas of their life (request for staff to attend a Social Services Review)
• before and after solution focused scales e.g.’ how confident do you feel in
  managing this child’s behaviour’
• All of these are child as client focused – on the outcome for the child, apart from
  the last one, ‘how confident do you feel in managing this child’s behaviour.’

Discussion of the survey
Replies to this survey indicate that::
• psychologists have considered that measurement of consultation relates to
  outcome not impact
• psychologists in this county see the child as the client.
• school improvement was not considered in measurement of impact, although
  there are various school improvement projects underway where the clients are
  the staff and pupils.
• there was no mention of the impact (measuring consultee perceptions) except
  from one very newly qualified psychologist who emphasised the teacher
  ‘recognising his or her part’. The measures for the Tier 4 Project are given in
some detail because these exemplify client measures. In general, the psychologists' responses focused on these sorts of measures and on ensuring that recommended strategies were carried out. These strategies were recommended by the psychologist, and there was no mention of joint problem solving, collaborative work, or of establishing how the consultations were viewed by school staff – the consultees.

- informal comments were made by three psychologists on the time spent on such measures. These comments indicated that such time was at the cost of service to clients.
- Half of the psychologists surveyed did not return forms Those who did reply may not be representative for those reasons discussed earlier (section 2.5). One of the two seniors replied, three of the four specialists replied. Only one of the seven part time psychologists replied so that these replies cannot be assumed to reflect the opinions of the part time psychologists, nor is it clear why they did not reply. It may reflect a general lack of response to questions about practice, or the fact that not all psychologists are implementing consultation.

Appendix 1 shows a proposed raft of evaluation measures to form a coherent measurement of user satisfaction. The two overriding considerations are that,

- measuring consultation should be part of a coherent approach to measuring user satisfaction, in keeping with best value,
- and that measurement should appear as short and uncumbersome as possible to both psychologist and user. Page 18 shows the proposed measure (also in Appendix 2)
Consultation effectiveness measurement: draft proposal

These questions are taken from the drafts of other counties' measurement proposals: **outcome** measures what may have happened for the pupils or staff (e.g. move downwards on SEN register, greater willingness to read, school attendance, parental support, behaviour policy, more focused IEPs), and **impact** is the change in perceptions, understanding or confidence on the part of staff usually as a result of consultation (e.g. how has the consultation increased confidence or feelings of self efficacy, control over the situation, understanding, liaison with parents)

In the first year select three pupils to follow up outcomes, and three members of staff for measurement of outcome and impact. You may want to use the rating scale on the next page to measure change.( rating scale is from county S appendix 3)

**Outcome Measures: examples of questions**

How is x getting on?

When we discussed him we suggested - - - - what did you try?

How well has it worked?

What changes have you noticed? What changes have other people noticed?
Impact measures

How much impact do you think the consultation had on how staff viewed the situation?

What was said in the consultation which was particularly useful?

Overall, how would you rate the value of the consultation with the educational psychologist

<table>
<thead>
<tr>
<th>Very useful</th>
<th>Not at all useful</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 9 8 7 6 5 4 3 2 1</td>
<td></td>
</tr>
</tbody>
</table>
4.0 Recommendations on measurement of outcome for County Y:

(see appendices 1,2,3,4)

In the light of these responses, how can the service for County Y move to measuring effectiveness of consultation? The current system in Y is piecemeal, having grown in an adhoc way and measurements are repeated, or omitted. At present there is no measurement of outcome and impact of consultation, nor is there any attempt at measuring the user satisfaction from what psychologists in the survey identified as the client group – the students. A rationalisation of all the current data collection might be used to fit the measurement of outcomes of consultation into a best value approach. Such measurement would be one indicator of value. This is already underway in one service (Hull Educational Psychology Service has recently circulated their approach, July 2000).

• The measure should be part of a coherent system of service evaluation reflecting outcome and user satisfaction
• It should have face validity, that is, appear to measure what it does measure.
• Psychologists and users should feel that it is an efficient use of time

For these reasons a single sheet of questions is proposed with one rating scale (appendix 2). Unlike County S it is proposed that this is filled in on one occasion during a follow up consultation. Responses will give qualitative data with one rating scale yielding a quantitative measure.²

² The service already gathers quantitative data 6 times during the year on visits, reviews, and time spent with children at various stages of the Code of Practice.
The service should:

- ascertain what proportion of psychologists feel that consultation is implemented and to what extent in their schools
- discuss these responses (anonymised) as part of a county service meeting
- discuss the proposed consultation effectiveness form and decide whether they wish to trial that or would prefer a system similar to the one proposed by County S (appendix 3). The service may wish to dedicate some time to training/professional development activities on consultation and school improvement issues and consider the findings of Hughes et al (1997) and Busse et al (1999) as well as the measurements being trialled in other authorities in the UK.
- undertake a pilot study of a measurement instrument for consultation impact and outcome over the course of the next two to three terms; the study should consider how other forms of measurement fit with this, such as annual school feedback on the educational psychology service, and parental questionnaires on the service.
- adopt a more coherent approach to evaluation combining feedback from schools, from parents and students, and from consultation, to be completed at a specific time during the school year. At that time psychologists could also complete evaluation of their continuing professional development.
- set up a small working group to consider how to seek feedback from students

There are two approaches to evaluation – a ‘professional’ and a managerial. (see Webster and Hoyle, 2000 for discussion of new professionalism). Measurement of
outcomes can satisfy both but individual psychologists will place different emphasis on these. Appendix 2 shows the measurement as part of one evaluative system. The school as service user has the school annual feedback (appendix 4), which gives a clear account of work with the school and intended outcomes, and invites school feedback on the past year and future directions. User satisfaction of parents is measured with parental feedback questionnaires, and the professional development log allows each psychologist to consider professional development in conjunction with these feedback measures.

---

3 A recent service meeting has evaluated where the service is in relation to consultative practice.
4 Professional relates to how each psychologist measures his/her own service to the client and professional development: a managerial approach considers the service as a whole, how it meets the aims of the various development plans, what the best value performance indicators should be.
5.0 Measurement of consultation outcome as part of service evaluation: future directions

In the last ten years there has been emphasis on competition to provide what were formerly local authority services. Psychology services, like other services, could be provided in a different way, perhaps contracted out. In order to judge whether a service is providing value and efficiency for money, various indices are used to show that the service is performing well and providing good value. The new duty of best value from April 2000 required local authorities to seek continuous improvement with regard to 'economy, efficiency, and effectiveness' (DETR 1999). Under the Local Government Act 1999 the government can prescribe BVPIs – best value performance indicators. Examples relevant to educational psychologists are the number of statements issued within 18 weeks, provision for children out of school and BV50 (best value 50), which refers to the education of looked after children; it requires an interface between the performance indicators for education and for social services. Performance indicators place a much greater emphasis on users, and on gaining feedback from users. One approach is to self evaluate using performance indicators; Thomas MacKay has applied this to four key areas of the psychological service in Scotland, :-


A review of the Psychological Services in Scotland by HMI identified the need for effective procedures to ensure a high quality service; a steering group was then set up to see how performance indicators could be used. Mackay applies the questions used by schools in How Good is our School (HMI Audit Unit 1996) to the psychological services in each of the above areas.
How are we doing?

How do we know?

What are we going to do now?

For each of the four areas four levels were identified, from excellent to unsatisfactory and performance indicators are organised under each area with specific themes. For example, Key area 4 is Service Delivery, and one theme is quality of research and development opportunities offered by the service to educational establishments and other users. The question is

How good is the quality of research and development?

Can this approach be applied to consultation and could it be used to measure the effectiveness of consultation? MacKay asks:

How good is the range of consultation and advice provided?

(MacKay, 1999,p18)

He suggests items to look for:

- a wide range of appropriate consultation and advice provided to all service users
- the service advises and supports educational management in relation to children and young people with special needs and to social inclusion and raising attainment
- consultation and advice provided to educational establishments as an integral part of core services
- consultation and advice relevant to the context and aimed at empowering service users in developing action plans and evaluating outcomes

The last item in particular refers to the quality of consultation. County Y can ask, as part of a best value approach

*How good are we now? (at consultation?)*

*How do we know?*

*What are we going to do now?*

### 5.1 Future directions for the evaluation of consultation outcome

Psychologists need to evaluate the service they provide, and if it is a consultative service then the outcomes of consultation will be part of this evaluation. There are different ways in which psychologists can evaluate their practice; the professional practice of consultative educational psychology requires an evaluation of the outcomes of consultation, managers require indicators to support comparative 'best value' type appraisal.

- It will be helpful to identify clearly impact and outcome measures particularly as it is unclear how the impact of consultation relates to outcome (Busse et al 1999). In the single county survey above, educational psychologists who responded viewed measurement as an outcome for the child.
• Different kinds of measurement may be appropriate for different types of work; this is clear in the list of measures specified above for the Tier 4 project, and it would be helpful to identify what will be measured at the start, to specify desirable outcomes.

• Measurement of consultation effects should be part of the overall evaluation of the service given to users and so should include school staff, parental, and child feedback as well as feedback from other agencies.

Finally, consultation is an iterative process; in consultation a working hypothesis is explored and strategies identified which fit with that hypothesis (Gutkin and Curtiss 1988). If the strategies do not work then staff and educational psychologist return to the drawing board in the next consultation. Any measurement of consultation outcomes should recognise that problem solving is a process; sometimes the process will be lengthy.
References


HMI Audit Unit (1996) *How good is our school?* Edinburgh: Scottish Office Education and Industry Department


James, W., (1890) *Principles of Psychology*, New York: Holt


Verkasalo M, Tuomivaara P, Marjaana Lindeman, 1996 *Fifteen year old pupils and their teachers values, and their beliefs about the values of an ideal pupil* in Educational Psychology Vol 16, No 1 March 1996 35-47

Watson J.B. (1919) *Psychology from the standpoint of a behaviourist* Philadelphia: J.B. Lippincott


Appendix 1: Proposed self evaluation of the educational psychology service in county Y

1. Trial a consultation feedback form (CEF) to be used with selected schools to measure perceived effectiveness of consultation – eventually to be used with all schools over a two year period. Suggested questions are in appendix 3

2. The school review of the service
   - what worked well
   - what difficulties were there,
   - what would you like to change
   - what would you like to keep
   - outline of support and intended outcomes for schools of concern, of serious weakness, or on special measures (appendix 4)

3. Parental feedback measures either by phone survey or forms issued to 5 families for each EP per year – families to be chosen from the data base randomly by senior. These should include families of children at every stage of the Code of Practice. EPs should know which families are selected in case there are family factors which make it inappropriate to seek feedback

4. A pupil feedback form for older students to be piloted – one EP to be responsible and to start with secondary age statemented pupils

5. The completing of the professional log demonstrating continuing professional development

Quantitative data could be returned at the same time detailing statistical returns for school visits, reviews, and pupil contacts.
Appendix 2: Consultation effectiveness measurement: draft proposal

These questions are taken from the drafts of other counties' measurement proposals: outcome measures what may have happened for the pupils or staff (e.g. move downwards on SEN register, greater willingness to read, school attendance, parental support, behaviour policy, more focused IEPs), and impact is the change in perceptions, understanding or confidence on the part of staff usually as a result of consultation (e.g. how has the consultation increased confidence or feelings of self efficacy, control over the situation, understanding, liaison with parents).

In the first year select three pupils to follow up outcomes, and three members of staff for measurement of outcome and impact. You may want to use the rating scale on the next page to measure change.

Outcome Measures: examples of questions
How is x getting on?

When we discussed him we suggested - - - what did you try?

How well has it worked?

What changes have you noticed?

What changes have other people noticed?

Impact Measures
How much impact do you think the consultation had on how staff viewed the situation?

What was said in the consultation which was particularly useful?

Overall, how would you rate the value of the consultation with the educational psychologist

<table>
<thead>
<tr>
<th>Very useful</th>
<th>Not at all useful</th>
</tr>
</thead>
<tbody>
<tr>
<td>10  9  8  7  6  5  4  3</td>
<td>2  1</td>
</tr>
</tbody>
</table>
## Appendix 3: Measuring change

What is your professional view of this pupil's performance **now** in any of the following areas? Please rate it by circling the appropriate number below.

<table>
<thead>
<tr>
<th>Area</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>not at all</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Learning</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>very satisfied satisfied</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>not at all</td>
</tr>
<tr>
<td><strong>Class behaviour</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>very satisfied satisfied</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>not at all</td>
</tr>
<tr>
<td><strong>Behaviour outside class</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>very satisfied satisfied</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>not at all</td>
</tr>
<tr>
<td><strong>Interaction with peers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>very satisfied satisfied</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>not at all</td>
</tr>
<tr>
<td><strong>Interaction with staff</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>very satisfied satisfied</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>not at all</td>
</tr>
<tr>
<td><strong>Self esteem</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>very satisfied satisfied</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>not at all</td>
</tr>
</tbody>
</table>

What would you hope your professional view of this pupil's performance will be six weeks after consultation?

<table>
<thead>
<tr>
<th>Area</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>not at all</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Learning</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>very satisfied satisfied</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>not at all</td>
</tr>
<tr>
<td><strong>Class behaviour</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>very satisfied satisfied</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>not at all</td>
</tr>
<tr>
<td><strong>Behaviour outside class</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>very satisfied satisfied</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>not at all</td>
</tr>
<tr>
<td><strong>Interaction with peers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>very satisfied satisfied</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>not at all</td>
</tr>
</tbody>
</table>
Interaction with staff
very satisfied 6 5 4 3 2 1 not at all satisfied

Self esteem
very satisfied 6 5 4 3 2 1 not at all satisfied

How confident do you feel about being able to manage the problem at the moment?
very confident 6 5 4 3 2 1 not confident

At our consultation last term you rated C---'s performance like this (show previous rating scale). Would you circle the number which best describes him now?

Learning
very satisfied 6 5 4 3 2 1 not at all satisfied

Class behaviour
very satisfied 6 5 4 3 2 1 not at all satisfied

Behaviour outside class
very satisfied 6 5 4 3 2 1 not at all satisfied

Interaction with peers
very satisfied 6 5 4 3 2 1 not at all satisfied

Interaction with staff
very satisfied 6 5 4 3 2 1 not at all satisfied

Self esteem
very satisfied 6 5 4 3 2 1 not at all satisfied

How confident do you feel about being able to manage the problem at the moment?
very confident 6 5 4 3 2 1 not confident

Much of this is adapted from a draft scheme from S--- EPS
## Appendix 4: School Feedback for the Educational Psychology Service


<table>
<thead>
<tr>
<th>Name of School:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>EP:</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Record of visits to schools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dates</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outline of support and intended outcomes (EP fills in)</th>
</tr>
</thead>
</table>

**School feedback (School fills in. Please use back of sheet if necessary)**

What has worked well?

If difficulties have arisen what have they been and how could they be solved?

What would you like your EP to continue doing?

How else would you like your EP to work with your school? .
Professional Practice Assignment

Submitted in part fulfillment of the requirements for the Continuing Professional Development Doctorate in Educational Psychology (DEdPsy)

Name of Course Member: Deborah King

Title of Assignment:

The Numeracy Strategy or recovery programmes for pupils with difficulties in numeracy.

Assignment number 3

Core Curriculum area to which this assignment topic relates:

Assessment and Intervention
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0 Introduction</td>
<td>95</td>
</tr>
<tr>
<td>1.1 The daily maths lesson for pupils with difficulties, and educational psychology</td>
<td>95</td>
</tr>
<tr>
<td>1.2 Numerosity and how it may relate to other skills</td>
<td>96</td>
</tr>
<tr>
<td>1.3 The Numeracy Strategy</td>
<td>101</td>
</tr>
<tr>
<td>1.3.1 Background</td>
<td>102</td>
</tr>
<tr>
<td>Practice in recall of numeracy facts</td>
<td>102</td>
</tr>
<tr>
<td>1.4 Mathematics recovery programmes</td>
<td>104</td>
</tr>
<tr>
<td>1.4.1 The Mathematics Recovery Programme</td>
<td>104</td>
</tr>
<tr>
<td>1.4.2 The Numerosity Recovery Programme</td>
<td>106</td>
</tr>
<tr>
<td>2.0 Case studies of pupils with some difficulties in numerosity</td>
<td>108</td>
</tr>
<tr>
<td>2.0.1 Comment</td>
<td>110</td>
</tr>
<tr>
<td>2.1 Differentiation in the Numeracy Strategy</td>
<td>111</td>
</tr>
<tr>
<td>2.2 The Mathematics Recovery Programme</td>
<td>114</td>
</tr>
<tr>
<td>2.3 The Numeracy Recovery Programme</td>
<td>119</td>
</tr>
<tr>
<td>2.3 Pedagogy and numeracy</td>
<td>120</td>
</tr>
<tr>
<td>3.0 Conclusions</td>
<td>123</td>
</tr>
<tr>
<td>3.1 Inclusion, differentiation, and practice</td>
<td>123</td>
</tr>
<tr>
<td>3.2 Maths and the educational psychologist</td>
<td>125</td>
</tr>
<tr>
<td>References</td>
<td>126</td>
</tr>
</tbody>
</table>
1.0 Introduction

1.1 The daily maths lesson for pupils with difficulties, and educational psychology

This assignment discusses the teaching of pupils who have difficulties in maths. It considers the Numeracy Strategy, introduced in 1999 (Department for Education and Employment, 1999) a nationwide programme for the teaching of maths, designed for all pupils in mainstream education. It also considers recovery approaches which are designed specifically for pupils with difficulties, the Mathematics Recovery programme, (Wright, Martland and Stafford, 2000) used in parts of Australia and the U.S, and the Numeracy Recovery programme (Dowker, 2000). Two pupils with marked difficulties in number have been selected as case studies, to illustrate both the extent of the various difficulties in the mainstream classroom, and to highlight some of the difficulties in adapting the Numeracy Strategy for pupils who are significantly behind their peers in numeracy skills. It will be argued that there are benefits from the emphasis on individual assessment and intervention in the Mathematics Recovery approach which are not offered by the Numeracy Strategy.

It has been difficult to find debate within the field of educational psychology on the teaching of maths to pupils with special educational needs; in general, educational psychologists appear to concern themselves less with children's learning of number than with other areas of school life - children's behaviour, self esteem, friendships, reading and spelling skills figure more widely in the educational psychology journals such as Educational and Child Psychology or Educational Psychology in Practice. There appears to be less emphasis on cognitive psychology and its application to
learning, especially the acquisition of mathematical skills and in particular, there
does not appear to have been much debate about the pedagogy of numeracy and
special educational needs. However, the assignment will discuss a paper by
Brahms and Norwich (2001) on pedagogical approaches to teaching numeracy to
those with special educational needs; it has argued for continua of teaching
approaches rather than recovery programmes.

The next section will consider some recent research on numeracy and its
relationships to other skills so that the two case studies, and the pedagogy of the
Numeracy Strategy and the Mathematics Recovery programme, can be considered
against this background. Although the Numeracy Strategy covers the mathematics
curriculum, including Shape and Space (geometrical skills) this assignment will
focus on the numeracy element, or numerosity, discussed in the following section.

1.2 Numerosity and how it may relate to other skills

Numerosity, or 'numerousness' is a property of an array, the number, or the amount
of. However, it is more than just knowing the number of a group of objects; Nunes
(1998) defined it as the awareness of the relatedness of numbers rather than their
absolute value. Brannon and Van de Walle (Brannon and Van de Walle, 2001)
pursued this knowledge of relatedness by pointing out two essential aspects of
number; one aspect is cardinality, which is the idea of one to one correspondence,
and the other is the ability to represent relations between numbers – that four is
greater than two. The word numerosity encompasses both aspects, cardinality and
the knowledge of relationships between numbers (Butterworth 1999) and is used in
this sense in this assignment.
It has been argued recently that humans have an innate capability to deal with numbers, a common structure predisposed to recognise the numerosness of sets. Dehaene, in particular, has claimed that there is a modular knowledge domain for number and that this is independent of linguistic ability or 'general intelligence' (Dehaene, 1997, Dehaene, Dehaene-Lambertz and Cohen, 1998). Dehaene quoted evidence such as the ability to subitize, experimental work with infants, and crosscultural studies. Subitizing was the ability to name the number of small arrays of dots or objects – generally up to 4 and sometimes up to 7, although the time taken increased markedly for arrays greater than 4 (Dehaene, 1996). Work with infants using habituation has shown that they were surprised by displays which violated the rules of number. The most famous of these was Wyn's experiment (Wyn, 1992) although Brannon and Van de Walle have raised the possibility that infants might have been using a limited ability to keep track of individual entities. If we accept Nunes' emphasis that numerosity is the relationship between numbers and the realisation that numbers are not absolute, it is more difficult to say that the infant studies demonstrate this. Less controversially, cross cultural studies have shown a large number of counting systems existed in the past and still exist across cultures. Butterworth has expanded on this (Butterworth 1999), and furthermore, an innate predisposition for learning some number concepts appeared to be supported by evidence from animal studies (Brannon and Van de Walle, 2001). Such evidence has been used to argue for a common structure predisposed to recognise numerosness or numerosity of sets. Both Dehaene and Butterworth have argued that there may be an innate ability to extract numerosity as a salient and useful fact from the environment (Dehaene 1996, Butterworth, 1999).

The relationship existing between the cognitive concepts of numeracy and other aspects of cognitive functioning has also been explored by the study of adults with
acquired difficulties in numerosity after brain trauma. Dehaene and Butterworth discussed acquired difficulties with number (Dehaene 1996, Butterworth 1999). Case studies may be used to illustrate the processing of numerical problems; adults with acquired difficulties are particularly interesting in the abilities they have retained and lost. Both Dehaene and Butterworth discussed apparently anomalous evidence when a patient had been unable to carry out a simple addition but could multiply. Such case studies may throw light on models of numerical processing, and in the same way, close study of what a child finds most difficult might help to establish a model of developmental progress in numeracy.

From the approaches of Dehaene and Butterworth the following questions are relevant to the learning of numerosity;

• how far is numerosity independent of other domains? Dehaene claimed that this domain relating to numerosity was independent of linguistic ability (Dehaene, 1997; Dehaene-Lambertz and Cohen 1998). How far does the development of numerosity depend on the development of language?

• how much depends on innate preparedness and on environment, particularly the exposure to a mathematical environment. Interaction may exist between an innate preparedness to construct a body of numerical skills and the environment; is there a stage at which the child is most prepared for mathematical input – a sensitive or even critical stage?

• could it be assumed that there is a continuum in numerosity; can a child be born with a greater predisposition to numerosity than another?

Although all of these questions relate to the learning and teaching of numerosity, only the first of these can be discussed briefly, because it is particularly relevant to the discussion of the two case studies below. What difficulties are likely to co-exist
with difficulties in numerosity? The complexity of the relationship existing between numerosity and other cognitive skills is exemplified by the exploration of Brannon and Van de Walle into the links between language and maths (Brannon and Van de Walle, 2001). They discussed three positions with regard to the relationship between language and the ability to make ordinal numerical comparisons; a 'strong' language hypothesis placed the acquisition of language as a necessary prerequisite for making numerical comparisons; a 'weak' language hypothesis indicated that language was necessary for larger sets; an 'irrelevant' language hypothesis was that children had some sort of preverbal numerosity existing without reference to language skills. Their work showed that children without any verbal numerical knowledge could not make judgements about which of two trays held the bigger number of objects and this is an indication of the part that language may play a part in the acquisition of numeracy concepts. Nunes also emphasised the close link with language in that children were found to have difficulty in naming the operation they had used to obtain the right answer (Nunes 1998) although, it might be argued that this is an output difficulty rather than a processing difficulty. A problem in separating maths from language is that language is most usually the medium through which numerosity is judged.

While acknowledging the likely relationship between numeracy and language, Nunes has also argued for the effect teaching maths has on the thought processes of the child:

‘mathematics is above all other things, a source of models for thinking’

Nunes, 1998 , 21
She maintained that numeracy provided ways of tackling problems, and frameworks for thinking, and pointed out the development of various aspects of numerosity following work with young children showing common errors in their ability to perceive the relationships between numbers (Nunes, 1998).
1.3 **The Numeracy Strategy**

1.3.1 **Background**

Besides the body of research indicating a possible innate predisposition to think mathematically, the background to the introduction of the Numeracy Strategy is one of concern at achievement compared with other countries. International comparisons show large cross cultural variation. While the effect of a culture bias in the test must be taken into account Butterworth (1999) points out the differences between Singaporean and Iranian children on the same maths test. The mean score for the Iranian children is equal to the poorest 5% of the Singaporean (TIMS 1996). Butterworth argues that this difference is due to culture because the older the children the bigger the gap becomes. Surveys of national and gender differences make interesting reading (see TIMS 1996-7) although variations are probably due to a number of factors. This concern over maths teaching and learning was not new. The Cockroft Report (DES 1982) drew attention to the lack of understanding often accompanying knowledge of formal maths rules. Many children could not apply the correct maths procedures to real life problems.

In the last ten years there has been an effort to make objectives explicit and lessons more formally structured, with an increasing emphasis on more prescriptive whole class teaching. The Literacy Strategy (1997) and the Numeracy Strategy (1999) have formalised objectives and content and are designed for all children in mainstream schools with an emphasis on whole class teaching and group activities during a prescribed time set aside daily. The Numeracy Strategy covers Reception through to year 6, and the recently introduced Early Learning Goals (DfEE 2001) for nursery and the first year in school have goals in line with the Numeracy Strategy. In terms of the Numeracy Strategy as an intervention it could be seen as both a macro
intervention to raise mathematical standards in all schools in England and Wales - and a micro intervention for each school and class; there are targets within the Strategy relevant to all children in mainstream schools. The Introduction details assessment, planning and differentiation. Additionally, the Numeracy Strategy and Standard Assessment Tests were designed to make it easier to identify those children having difficulties with number and lends itself to easily identified and measurable targets for action plans. As well as the emphasis on clear objectives and careful planning, there is a guide to good classroom practice in clear unambiguous language (The Numeracy Strategy 1999) and clear and precise lesson plans with explicitly stated objectives.

1.3.2 Practice in recall of numeracy facts

Ofsted provided reviews of recent research (Askew and William 1995) and one example of this underpinning from research is the emphasis on practice and memorising of arithmetical facts - number bonds and tables (Sharma, 1990, Nunes and Bryant, 1996, Butterworth, 1999). In most industrialised society time taken to solve arithmetical problems decreases with age and at ten it may be half what it is at eight. This speeding up is due to practice and memory - learning for example, that in adding any order is fine, and memorising number bonds so that retrieval is automatic rather than relying on calculation. This difference in speed is still there for adults and faster speed goes with a greater amount of retrieval from memory rather than working out. Another illustration of practice is that the fastest people to work out arithmetical calculations may not be mathematical prodigies but skilled cashiers; Butterworth has detailed the part practice plays in displays of mathematical ability (Butterworth 1999) indicating it was an important variable, as was the amount of number facts stored in memory. Sharma (1990) agreed with this when discussing clinical work with children with learning difficulties:
'most of these children have - - slow or no recall of arithmetic facts'
Sharma (1990) p3

This was reinforced by both Butterworth (1999) and Nunes and Bryant (1996); speed of calculation comes from a retrieval of known mathematical facts and this was mentioned in an Ofsted report on schools which have been subject to special measures:

'in sessions devoted to mental maths in a primary school, beat the clock is a popular strategy to sharpen the pupils' calculating skills. They have improved their speed of recall so significantly that the staff are considering reducing the frequency of these sessions because they are no longer as necessary for checking that the pupils know the basic number facts' paragraph 41 (Office for Standards in Education 1998).

This appeared to be one of the main planks of the Numeracy Strategy, the learning and recall of number facts: it also emphasised as it does, the need for knowledge of the language of maths, for rapid recall, group work, understanding, the use of visual aids such as number lines and squares, and of other visual and concrete apparatus - a bonus for those whose learning style may be visual and concrete (see Riding and Rayner 1998 for further discussion of learning and cognitive style). It might be argued that the children most in need of this very structured approach to learning and the recommended strategies and aids to explanation are those children who have special educational needs. In other words the children who have some difficulties - but not marked difficulty with number - might benefit from this structured and linear approach and emphasis on practice and the recall of number facts. However, the Numeracy Strategy may not have been designed for those children
who have not the understanding of the numerosity underlying the recall of number facts:

'It (the Numeracy Strategy) does not give sufficient detail, nor provide the less confident teacher with support firmly grounded in theory, to help the less able children - -'

Wright, Martland, and Stafford 2001, 18

This comment has, however, not come from a disinterested group, but from those who have implemented a recovery, or what might more accurately be termed a restart approach, described below.

1.4 Mathematics recovery programmes

1.4.1 The Mathematics Recovery Programme

In contrast to an emphasis on the recall of number facts, and what they term as a 'top down' mathematicians approach to learning the concepts of numerosity, Richards, Steffe and von Glaserfeld have adopted a constructivist approach; they emphasised the part played by the child in constructing his or her own learning. Their approach to number teaching was, they claimed, bottom up, rather than the traditional top down approach stemming from mathematicians. Their curriculum design started with the learner while they believed mathematicians had designed the curriculum, starting from the idea of mathematics,

'disembodied from human experience'

Steffe, in Steffe and Wood, eds., 1990, 390
They have also pointed out that the number system has developed over thousands of years (Steffe 1992a, Steffe and Cobb 1988, Steffe et al 1983, Steffe and Wood, 1990). There may not be a necessary contradiction between this and the belief in an innate predisposition; a predisposition may exist towards constructing numerical concepts and systems but the formal system used today has evolved over time. However, the constructivist position of Steffe and his colleagues, that the child constructs her learning from the environment, has led them to place emphasis on how children learn to count, and a close scrutiny of what children do when presented with counting problems has been used as the basis for the Mathematics Recovery (MR) programme in Australia (Wright et al 2000). – Like Piaget, (Piaget, 1952) they saw the development of number as a gradual emergence, governed by whichever stage of development the child is at. Their observational approach, followed by close analysis and the postulation of various counting types or stages was reminiscent of a Piagetian approach and the authors acknowledged his influence in their recommendations for the design of a mathematics curriculum which emphasised a mathematics learning environment for assimilation (Steffe and Wood, 1990).

They characterised approaches to mathematics teaching since the 1930's as top down approaches – derived from the subject matter rather than from the learner. Furthermore Steffe believed that the psychology of teaching maths has been behaviourist since the 1930's, based on a stimulus response methodology introduced by Thorndike (Thorndike 1911). From this came the drill and practice methods and so he postulated, came the difficulties with maths problems since children had not the understanding to apply what they had learned in a behaviourist routine.
There appear to be some difficulties with this; firstly, it is surprising that Cockroft identified the same difficulty in 1982; he too argued pupils could not apply school mathematics to real life problems, following, it could be argued, a decade of Piagetian discovery method in teaching. Secondly, it would be difficult to use a truly constructivist approach which prescribed stages or counting types through a set assessment. The MR programme diagnosed which stage a child is at through a set assessment and then delivered a tailored programme. There was an assumption that children’s cognitive processes could be grouped and a common programme applied to each of these groups. As Piaget has been criticised for the invariance of his stages, it may be claimed that this approach also tries to place children at stages. It is noticeable that Steffe called these counting types rather than stages but the Mathematics Recovery Programme assumed they were stages which were sequential.

1.4.2 The Numerosity Recovery Programme

In Oxford, a similar intervention programme, Numeracy Recovery, has been carried out by Dowker (Dowker, 2001). Dowker pointed out the success of various reading recovery programmes (Sylva and Hurry, 1995) and argued for a Numeracy Recovery programme for 6 and 7 year olds before they developed expectations of failure, and the wrong use of strategies, as has happened in the case of NR. Dowker, in common with the Numeracy Strategy, emphasised the part played in numerosity by the learning and retention of number facts. While it is similar to the MR programme in that it starts with assessment, the stages were differently derived from the principles involved in counting (Greeno, Riley and Gelman 1984). The Numeracy Recovery programme started with assessment on 8 components but had a much less intensive intervention - only one half hour intervention weekly, from the
class teacher for 30 weeks. It is uncertain how far Dowker's assessment resembled the MRP as the description of the MRP by Wright detailed the stages the assessment was based upon, but not the actual assessment, possibly for commercial reasons.

The differences in the underlying philosophy of the Numeracy Strategy and the Mathematics Recovery Programme have led to their different approaches to pupils who have difficulties with numerosity: the Numeracy Strategy adopts a common curriculum with differentiation, and the Recovery Programmes start with an attempt to see what stage the child's learning is at, with individual assessment and a tailored intervention. These approaches will be considered in the next section in relation to two pupils with quite marked difficulties in maths.
2.0 Case studies of pupils with difficulties in numerosity: approaches to their learning

These two case studies were selected on the basis of the prime difficulty being with number. Appendices 1 and 2 show parts of the numeracy strategy for years 3 and 6. The pupils are in years 6 and 7.

Case study 1

N.R. was first referred at the age of 8.3 and although her reading was age appropriate, she sometimes had some difficulties in interpreting what she was reading. However, she was referred because her difficulties with the smallest addition and subtraction were evident. These placed her at about the first centile. The other quite obvious areas of difficulty showed themselves in the science curriculum - she could not sort or classify - for example things which were manmade, or things which used electricity. Early interventions focused on number skills making sense - so she was asked to set the table and to count out the cutlery for four. Her mother said her success in this was variable and she would often have the wrong number of knives or forks. Calculations with money up to 10 pence were too difficult.

In middle school intervention focused on persuading N.R. to ask for help. By the age of 9 she had become increasingly anxious about number work and science and although she would copy work from the board and text book carefully she is still unable to carry out quite simple calculations. At the age of 8.9 if asked to add 1 to 99 she would draw 99 strokes leave a gap then add 2 more strokes.

\[ 11111111111111111111111111111111 \] (99 of these) then she draws another two 1 1. She then counted all the strokes from the start and arrived at 102.

When she was two years older - 10 years old - she wrote down

\[ \begin{align*}
99 \\
2 \\
\end{align*} \]

\[ = 78 \]

She explained she had added the two 9s to get 18 and then took the 2 from 9. She still does not understand the relationship of 99 to 100. She has knowledge of procedures but uses them wrongly together to produce an answer which isn't sensible. At 10 years and 8 months she adds 3 to 9 to make 10. She finds subtraction as difficult although probably no more so.

N.R. is sociable, good at games and P.E., but is increasingly anxious about her difficulties with number. Staff at her middle school have suggested special education, using the argument that special school would focus on maths 'for life skills' and build up her confidence.
What interventions would be helpful now? Her difficulties are with number work - yet she copes with shape and space. She does not understand verbally presented information particularly in maths and science. It might appear that in some areas she has not reached the concrete operational stage. Her difficulties are with invariance, transitivity and seriation. The intervention needs to address these - and so it would be helpful to allow her to work with a younger group for numeracy - perhaps at key stage 1. Staff at her middle school have said they cannot differentiate to such an extent. She needs to gain confidence - yet any intervention underlines her difficulties. She has quite enough social sense to realise her difficulties set her apart in certain areas. She has become increasingly anxious and tearful.

Case study 2

C is noticeably better at shape and space work than number work. Although he has learnt his tables and has some strategies for mental calculations, when put on the spot he appears to have genuine difficulties in sequencing the tables and ordering his thoughts. He doesn't like appearing to be unable to do something so applies all sorts of avoidance/covering up strategies. (Report from maths teacher June 1999)

In February 1999, when he was 12 years and 3 months, I asked CW to do some work from the Wechsler Objective Numeracy Dimensions (WOND 1996). To take away 9 from 12 he drew 12 strokes and then crossed out 3 and counted what was left - giving the right answer. To divide 16 by 2 he again drew 16 strokes and placed a line in roughly where he thought the middle was. Counting the strokes on one side gave him 9. Of course, as the numbers grow larger these methods tend to give the wrong answer and take a great deal of time.

C.W. also has difficulties in understanding of information given verbally - and so he reads slowly and takes considerable time to process what he has read. He tends to forget sequences of instructions and so forgets how to work through a mathematical procedure such as addition of large numbers. He takes a considerable time to think about answers to questions. He dislikes asking for help. He is popular both with the staff and his peer group. His social skills are good and his company sought after - but is increasingly irritated by what he sees as unreasonable homework and study club demands - interventions which assume a catch up is possible with greater input. The difficulties are those which Sharma (1990) discusses at some length in considering the types of number and maths difficulties which may coexist with other difficulties such as dyslexia. He misreads the question, does not remember which mathematical procedure to apply and forgets the correct sequence of quite basic number skills such as addition and subtraction. He often misreads or forgets to take account of the sign plus or minus. He works very slowly and avoids attending homework club. Staff find his behaviour is increasingly difficult in class although all find him pleasant on a one to one basis and his social skills are good. Increasingly in maths class he appears to lack attention and motivation and is perceived by staff as in danger of becoming disaffected.
2.0.1 Comment

For these two the difficulties are not specific - but numeracy is the prime reason for and focus of intervention. However, it is worth remembering that a Piagetian approach (Piaget 1952) might assume that with such difficulties in numeracy they might also have difficulty with logical operations - and so any education plan must consider Science, and perhaps Design Technology. Sharma talks of a difficulty in conceptualization; both N.R. and C.W. do not see what might be a sensible answer - so estimation is difficult and answers may not 'be reasonable.' (Ofsted, 2000 )Both find it hard to choose the correct procedure for a given problem because there is not understanding of how to apply number to solve problems. In both cases there are difficulties with verbal understanding, and with the basic concepts and relationships of number - ones to tens and tens to hundreds. Of the two cases described above, NR appears to have a greater degree of difficulty with numbers and their relationship than did CW. NR has no idea of which numbers may be greater in a comparison of two numbers over 100, she could not subtract or divide, nor use object number correspondence in setting the table; that is, for her own family of 4, she could not work out that 4 knives and 4 forks would be needed.

Both know that they are likely to be wrong - and so both avoid situations which are part of interventions. N R has little confidence and may cry easily and is quite sure her answers are wrong. CW avoids situations where he has to deal with number - and is ambivalent about asking for help. Both N.R. and C.W. have larger problems than difficulty with number although their problems are less evident in other areas of the curriculum. They find it hard to answer questions about a passage.
The relationship between language and numeracy discussed by Brannon and van der Valle (2001) and by Nunes (1998) is particularly relevant to CW, whose difficulties coexisted with a recognised language difficulty. It is not possible to say whether a weak or strong language hypothesis would be supported because there is not sufficient data. However, he appears to have some visual strategies for division – to draw the required number of strokes and make a mark in the middle. He uses his spatial awareness to help him split the number. The ability to use imagery, pattern or spatial imagery was discussed by Presmeg (Presmeg 1997) who argued that Einstein was a visualiser and, in following children's progress through school, that the visualisers did not do so well as the verbalisers in maths, because maths teachers tend towards a verbal rather than a visual approach. She also argued that the use of imagery has been devalued in many classrooms.

2.1 Differentiation in the Numeracy Strategy

What will happen then for pupils - such as the two not unusual case studies - who at 9 or older are still struggling with the basic number concepts?. There may be effects on self concept and on social groupings for individual pupils who find themselves working with a younger age group, or for pupils with their own age group who have not understood the lesson. The Numeracy Strategy had a variety of recommendations in the introduction (The Numeracy Strategy, 1999, pp 18-24) which focused on differentiation. The Numeracy Strategy (1999) mentions up to four groups for activities and that staff should ensure:

'differentiation is manageable and centred around work common to all the pupils in a class with targeted, positive support to help those who have difficulties to keep up with their peers'

(DfEE, The Numeracy Strategy 1999 p.5)
For pupils with difficulties in maths the Numeracy Strategy was envisaged as being suitable using the same subject matter but with differentiation through graded activities for three or four groups in the activity session and in the questioning of the whole class session.

'Teachers can involve and support all pupils through the strategies for differentiation','

The National Numeracy Strategy, DfEEp22

The later addition of *Springboard, the Booster Pack, and Mathematics for Pupils with Difficulties* (DfEE, 2000 QCA, 2001) were aimed specifically at pupils experiencing difficulty but the first two were for pupils in that fourth group in the class, and the last for pupils with profound and multiple difficulties, often not in mainstream education. The additional materials are for children who may be on the early stages of the special educational needs register and insofar as the Green paper, *Excellence for all* emphasised the need to stop these children moving up the register (Green Paper 1997), such additions are useful but none was aimed at pupils, such as NR and CW, where there was such a gap between them and the class.

However, in early consultation on numeracy it was advised that differentiation was kept to a minimum (Askew and Wiliam, 1995 DfEE, 1997). In the light of concern about the British position in the national league tables for numeracy it was felt that the greater the degree of differentiation 'downwards' the slower the progress for the majority. Despite this, the lesson plans recognised that here may be as many as four groups, though early recommendations were that there were no more than two,
To differentiate effectively takes careful assessment and planning (DfEE, 1999) and, the author would argue, assessment of the level of the pupil's current understanding, and how she will move on to the next stage. This is time consuming and may be labour intensive, and for pupils working at a very different level from the class it is not always possible to differentiate the same material or even to work on the same number processes. The differentiation recommended in the Numeracy Strategy is unclear about whether the task is differentiated, or whether the understanding of the concept is differentiated or whether it is 'differentiation by outcome' when all have the same task but show differences in the way they tackle it or in the level of answer. The differentiation recommended by the Numeracy Strategy appeared to start with the task rather than the understanding of the pupil, in that the recommendation is that there is, regardless of understanding:

'work common to all the pupils in a class'


Differentiation, then, appeared to be focused on adaptation of the material and a glance at the appendices will show the difference between CW and NR and the programme designed for their age groups. The appendices 1-4 show the appropriate work for year 6, and for year 3; emphasis for year 6 is on long division, and such concepts as probability and ratio. NR and CW are working at or below the level for year 3 pupils where there is emphasis on multiplication and the relationship between multiplication and addition. Probability is introduced in year 3. In Vygotskian terms the level of understanding of NR and CW is not near enough to that of their age group to work on the same material even with the levels of differentiation recommended by the numeracy strategy (Vygotsky, 1978). Pragmatically it might be recommended that these 12 and 11 year old pupils work
with pupils 5 years younger than themselves but this runs counter to current inclusive thinking (Booth et al, 2001).

The interim evaluation of the Numeracy Strategy has indicated difficulties with differentiation, in that it can be time consuming and may not be effective:

‘the intention of reducing the gap between the attainment of the different groups is unlikely to be met’
Office for Standards in Education, 2000,10

The same interim report recognised that standard written methods:
‘are of no use to someone who applies them inaccurately and who cannot judge whether the answer is reasonable’
ibid, 10

The author would argue that differentiation is an unclear concept, with a variety of meanings and particularly unsuited to the teaching of numerosity where learning and understanding may be based on the understanding of previously learned concepts. For example, there would be little point teaching place value to hundreds if the child had not understood place value for units and tens.

2.2 The Mathematics Recovery Programme

On the other hand the Mathematics Recovery programme was based on the idea of establishing the stage of numeracy reached by the individual child and then tailoring a programme for that child. In 1983 after Steffe and his graduate students had completed a large scale observational study of young children tackling numeracy problems, they claimed that:
for the first time there is an explicit foundation from which to explore the steps children have to make to acquire whole number concepts and numerical skills'

(Richards, Steffe, and von Glasersfeld 1983 114)

'the counting types theory, however, does provide a model of the mathematical reality of the child, allowing us to understand the child’s numerical concepts and operations'

(ibid, 121)

Wright et al (2000) described the Australian Mathematics Recovery programme (MR), also based on the idea of mathematical stages of learning and emphasised the assessment and teaching element based on:

'a strong underpinning theory for children's math learning' (Wright et al p2)

Wright's approach was functional (Wright et al 2000). He pointed out that children start school at different levels of numerical knowledge which may represent a three year difference. The difference between children could increase to a 7 year gap after 10 years at school:

'the knowledge gap between low-attaining children and average or able children tends to increase over the course of their years at school'

Wright et al, 2000 p1

It is worth being a little wary of this; the concept of a seven year gap was presumably based on the expectations of the more able as well as those with
special needs. Of two sixteen year old pupils one might be performing at above the level expected of an eighteen year old (perhaps A level standard in England) and the other at the standard expected of an eleven year old.

Wright goes on to describe the Mathematics Recovery programme, a 10-15 week programme for low attainers of 6-8 year olds; they were given intensive individualised teaching based on a learning framework for assessment and teaching. Teachers were trained to adopt a special instructional approach through a professionally developed course and with ongoing support during the instruction. A great deal of importance was attached to assessment of each child using set counting tasks to establish what counting type was being used by the child. (Steffe appeared to avoid the word stage but it is not clear how 'type' differs from 'stage'). The stages for assessment were based on those of Steffe and Cobb (1988) and were:
Stage 0: emergent counters – cannot count a visible collection of objects

Stage 1: perceptual counting – can count what is visible but not screened objects and cannot count two collections.

Stage 2: figurative counting: count from 1 and can add collections of screened objects - uses representation rather than concrete displays of objects. Can push together two collections of visible counters and count them.

Stage 3: initial Number sequence counting on and counting down (NR, the first case study cannot do this)

Stage 4: intermediate number sequence: counting down to identify missing number (the missing subtrahend task):

I have 12 and then 8: how many were taken away?

Stage 5: facile number sequence: range of skills other than counting by ones to add or subtract. can use groups of numbers or deduce from other calculations.

Children identified by teachers as having difficulties with maths were assessed on 8 components of early numeracy. The programme assessed counting procedures, use of arithmetic symbolism, use of place value, understanding and solution of word problems, translation between concrete verbal and numerical format, use of derived fact strategies for calculation, arithmetical estimation, and memory for number facts. Video assessments were made of children on a range of tasks, assessments were interpreted methodically and in detail, and placed on a stage of mathematical development, and then taught according to a programme tailored to this stage of development. The programme is built on teaching techniques such as ‘microadjusting’, ‘modelling’ and ‘demonstration’. There was a bank of instructional settings and tasks and the teacher chose from these to suit child’s level – just
beyond the current level of functioning, in other words, in the zone of proximal development (Vygotsky, 1978)

This programme was intensive being typically daily individual teaching, four days weekly for a programme lasting between 7 and 20 weeks (the average was 13 weeks).

'\textit{a participant should be discontinued when they have reached a level at which they are likely to succeed with regular class room teaching}' - Wright et al 2000,152

The programme appeared to have theoretical underpinning provided by the studies carried out by Steffe and colleagues, and the detailed and thorough assessment should give a sound basis for tailored intervention. However, the account given by Wright and his co-authors was disappointing on the following grounds. There was insufficient discussion of the effect daily individual tuition from a teacher might have without the assessment and programme; the evaluation appeared insufficient to allow firm conclusions to be drawn; there was an assumption that all children would follow a similar pattern in the proposed counting stages, or types.

What might have been expected from daily individual tuition from a specialist teacher without the intervention materials and method? Evaluation could have taken this into account with a control group using other interventions. The description of evaluation by Wright and his colleagues was disappointing. Outcomes for the 1994-5 Australian cohort and the 1995-6 US cohort were detailed in the appendices but there was little statistical analysis of the results of 89 children in Australia and 91 in the USA. Wright and his colleagues detailed percentages of
children who moved up the described stages but a convincing analysis was omitted. To be sure of the effect of the Mathematics Recovery Programme analysis of participant children should have been compared with two other groups – a non participant control group under usual curriculum conditions and a non participant control group with a similar individual tuition schedule. The MR programme has been used in the UK (Wigan) and in parts of the US so it is hoped that long term evaluation using control groups will be available.

2.3 The Numeracy Recovery Programme

In the Numeracy Recovery programme designed by Dowker (Dowker 2000) there was also initial assessment of counting, addition and subtraction, and arithmetical symbolism – reading and writing numbers, place value – tens and units, word problems, and 'derived fact strategies'. This last related to being able to use the information:

\[ 23 + 44 = 67 \]

to tackle

\[ 23 + 45 = ? \]

She has also focused on the idea of giving a sensible answer, (Ofsted 2000 pointed out that children should be able to judge whether an answer is reasonable). Children were asked to estimate answers by rating the answers given to a problem from 'very good' to 'very silly'. Like the Numeracy Strategy she emphasised learning number facts; the retrieval of number facts was assessed, e.g. 3+3, 6+6. Dowker's assessment appeared to focus less specifically on counting, addition and subtraction than the MRP and she did give details of the interventions, unlike accounts of the MRP by Wright et al (2000). Additionally evaluation appeared at preliminary stages to be more although evaluation of the Oxford Numeracy
Recovery Project is not yet complete. Unlike the evaluation discussed by Wright she has used a control group but unfortunately it was not clear how this group was selected and matched with the experimental group. Evaluation additionally had been to retest with a subtest of the Wechsler Intelligence Scales (WISC) after only 6 months and the question of a possible practice effect arises. Scores on the BAS number test appeared to have risen from 94 to 98 and a Wilcoxon test showed a significant difference between groups. Unfortunately there is not sufficient detail on selection of control groups and in the statistical analysis to conclude firmly that the programme is effective but Dowker has stated that evaluation will take place over three years to see if gains are sustained, and that the intervention will be compared with other interventions (Dowker, 2001).

2.4 Pedagogy and numeracy

Both the Mathematics Recovery Programme and the Numeracy Recovery Programme have attempted evaluation, whatever the weaknesses in this evaluation have been. Norwich and Lewis have made the point that research into the pedagogy of maths has been weak; there has been -

little systematic research into teaching pupils with low mathematical attainment and specific mathematics learning difficulties by comparison with the teaching of literacy

(Norwich and Lewis, 2001 p321)

Surprisingly, they have discussed neither of these recovery programmes. Instead they argued against recovery type approaches but for what they have called continua of teaching approaches: this would be in adapting the curriculum to a greater extent for those with greater need. The authors questioned the sustained effectiveness of recovery approaches in reading compared to other teaching approaches. However, their discussion of maths interventions is brief, by
comparison, and focused on secondary age pupils. They argued for an increase in
the number and range of examples and the amount of explicit instruction.

Without explicit instructional priorities - - - pupils - - are less likely to master
fractions, decimals, ratios or proportions

(Ibid, p321)

Their discussion did not appear to offer much to teachers considering how best to
teach pupils with difficulties in maths. Firstly, Norwich and Lewis did not consider
pupils in mainstream with such difficulties as NR, nor even CW, although they
discussed the needs of pupils with profound and multiple difficulties later in the
same paper. Fractions, decimals, ratios and proportions are part of neither NR nor
CW's learning programmes at present. All of these concepts require specific and
detailed knowledge of the relations between numbers or numerosity and at present
neither NR not CW appears likely to progress to the type of mathematical
understanding required for these operations and concepts. Secondly, it was unclear
on what evidence the recovery approaches in numeracy were rejected. It would be
useful to know what evidence there is that repeated practice and instruction helps
pupils who have significant learning difficulties in maths. In the case of both CW and
NR who have had extra maths tuition both at school and at home, in the first
principles of addition and subtraction for 4 years, using extensive numbers of
examples and explicit instruction, they appear to have made little progress. To
follow the parallels suggested with those adults who have acquired mathematical
difficulties, there is little to suggest that such pedagogy would reinstate their
understanding of numerosity. Norwich and Lewis assumed that it is possible to
teach NR and CW ratio, or fractions. If we take the Dehaene position of a specific
domain for numerosity which may have been damaged why should we assume
basic understanding can be put in place by more and greater ranges of examples?
Finally, Norwich and Lewis did not mention individual assessment and analysis of difficulties; these are traditionally the area of the educational psychologist although within a problem solving, consultation type of service (for example, see Wagner 1995) there may be less emphasis on this element of educational psychology. Sharma has summed up the view for traditional assessment models:

'without meaningful diagnosis the remedial effort may be a wasted effort at worst and getting minimum result at best'

(Sharma 1990, p25)

Terms such as Sharma has used - diagnosis - and a concern with whether difficulties are specific or occur in conjunction with other difficulties, indicated a need for individual assessment, particularly for those with significant difficulties, possibly warranting statutory assessment.
3.0 Conclusions

3.1 Inclusion, differentiation, and practice

The formal structure and prescribed content will probably raise achievement for most children, but for some it will emphasise the difference in their ability. The numeracy strategy for their age group will be too difficult - even when differentiated as suggested - for children such as CW and NR. For very able children it may be too easy. The more formal structure of the maths lesson may make this more obvious than before. The case studies illustrate the weakness of what might almost be described as the mantra of differentiation which the Numeracy Strategy uses as its main plank to help pupils with a range of difficulties in numerosity. The Numeracy Strategy fails to tackle adequately the nature of the differentiation required for pupils with significant difficulties in numerosity.

‘An assumption underlying recent government documents is that the effective curriculum is broadly, a common curriculum for all pupils’

Norwich and Lewis, 2001, 313

Within the Numeracy Strategy there is appropriate material for NR and CW but it is within the curriculum recommended for pupils four years younger. It is unclear how presenting similar subject material to pupils at very different levels from the rest of the class can be useful for those pupils.

The interim evaluation of the Numeracy Strategy has indicated difficulties with differentiation, that it can be time consuming and may not be effective (Office for Standards in Education, 2000) and that the standard written methods are not useful for pupils such as the two above,
'who cannot judge whether the answer is reasonable'

Office for Standards in Education, 2000,10

The interim report went on to comment that the gaps in classes could be narrowed by setting recognising the problem but failing to give clear guidance on how to tackle it. What are the implications for LEA targets for inclusion? It may be argued that an increase in good practice will help inclusion - but the more firmly laid down and less flexible curriculum content and delivery pace are, the less room there is for accommodation of those whose abilities mean they lie outside the curve of normality. The interim report pointed to an increase in setting for maths but made no evaluation of this.

The use of practice, or more accurately recall of memory facts, is another strand in the Numeracy Strategy which raises questions for the two case studies; the research on the effect of recall of memory facts (for example, Sharma, 1990, Butterworth, 1998) has been for children and adults with understanding of how to apply the memory facts. For NR and CW there may be less benefit in knowing number bonds to 20 or 100 unless they can apply these, for example in but in conjunction with money, 100 pennies in one pound, or cents in a Euro. The assumption lying behind the benefit of quick recall of memory facts is the understanding of when these are needed.

This brings us to what it is about the recovery programmes discussed above which sounds good for learning. They assess the level of understanding of the child and draw up a programme for that child. Wang, 1990 described effective instruction as instruction based on the assessed capabilities of each learner. Recovery programmes appear, as Steffe claimed to be bottom up programmes, starting with
the cognition of the child. A top down programme starting with the mathematical concepts to be taught carries with it an assumption of universality — that there are mathematical truths common to all, yet it is acknowledged that:

‘there is no single unique subject called mathematics but rather that it is a constantly evolving set of tools ——. There is no single correct form of mathematics nor a single correct way of teaching it’
Hughes 1986, 183

3.2 Maths and educational psychology

Returning to the issue raised in the introduction, of the educational psychologists’ focus on reading and social relationships, and not on numerosity, the author would like to suggest greater involvement of educational psychologists in numerosity interventions and policy making. It would be useful both for the young people who have such needs and for reminding educational psychologists of the importance of applying the principles of cognitive psychology to numeracy.

At the individual level designing and setting up such interventions educational psychologists could be particularly useful in the evaluation of recovery programmes and other interventions, in ensuring that control groups were set up and progress of groups properly compared. One way to approach this would be by the setting up of an interdisciplinary group including research psychologists such as Wright and Dowker, who are working on recovery programmes, mathematics consultant teachers and advisors, and educational psychologists, to advise the government on interventions for pupils with special needs which include difficulties with number.
References


Booth T, Ainscow M., Black-Hawkins, Vaughan M., Shaw, L., Index for Inclusion: developing learning and participation in schools. Bristol, CSIE Centre for Studies on Inclusive Education


Cockroft W.H 1982 see Department of Education and Science 1982


Dowker, A., 2001 Numeracy recovery: a pilot scheme for early intervention with young children with numeracy difficulty in Support for Learning vol 16, no 1 p6-10


127


---

1 Keys and TIMS refer to the same study. It is entered twice because the survey is often referred to as the TIMS study rather than by the names of the authors
Norwich B., Lewis., A 2001 Mapping a Pedagogy for Special Educational Needs in 


129


Thorndike (1911) *Animal Intelligence* New York, Macmillan

TIMS Study: see Keys


Wang M. C. (1990) Learning characteristics of students with special needs and the provision of effective schooling in M.C.Wang, M.C. Reynolds, H.J. Wahlberg (eds)


There are more than 6 papers referring to dyslexia or literacy and one on maths and precision teaching


Twelve papers refer to reading but none to maths.
Professional Practice Assignment

Submitted in part fulfillment of the requirements for the Continuing Professional Development Doctorate in Educational Psychology (DEdPsy)

Name of Course Member: Deborah King

Title of Assignment:

Communication skills in the statutory process: the professional development of one educational psychologist.

Assignment number 4

Core Curriculum area to which this assignment topic relates:

Interpersonal Effectiveness
Contents

1.0 Introduction 134
   1.1 Rationale for focusing on the communication element in statutory reports 135

2.0 Literature review 138
   2.1 Selection of literature 138
   2.2 Statutory assessment and educational psychologists 139
   2.3 Appraisal of educational psychologists and 360 degree feedback 143

3.0 Reflexivity and 360 degree appraisal 149
   3.1 Interpersonal communication items in the 360 degree appraisal 152
   3.2 The validity of this feedback procedure as a measurement of development or change, and methodological issues 155
   3.3 360 degree appraisal and professional development of communication skills 157

4.0 A framework to integrate professional and managerial development 160

References 163

Appendix 1 166
Appendix 2 167
1.0 Introduction

This aim of this assignment is to consider firstly how the doctorate may have influenced one area of interpersonal effectiveness in the work of an educational psychologist, and secondly, the development of skills in that area of interpersonal effectiveness. Interpersonal effectiveness is part of the core curriculum in the training of educational psychologists (British Psychological Society, DECP Training Committee 1997). The focus will be on communication skills, the fourth key skill area of interpersonal effectiveness:

'communication skills: reports: quality of professional advice and how it could improve problem solving of complex human issues'

(1997, BPS, DECP Training committee)

This excerpt refers specifically to communication skills in reports and professional advice, part of the statutory assessment process. The communication skills needed in the process of statutory assessment will be discussed; how they can be measured to indicate the possible effects of professional development, and the growing emphasis on evidence based assessment as it relates to this area. In section 4, my professional development will be discussed with reference to the issues raised by two 360 degree appraisals in the context of professional outcomes (Webster and Hoyle, 2000). These two appraisals were carried out at the start and towards the finish of a professional doctorate; part of the intention was to measure the effect of this professional
development. As this assignment considers my interpersonal effectiveness, the first person will be used, instead of the more usual convention of the third person.

1.1 **Rationale for focusing on the communication element in statutory reports**

The process of preparing advice for a statement of special educational needs is a significant element in my work as a local education authority educational psychologist. The work related to the psychological advice for statutory assessment remains an important part of my overall workload, taking up between 25% and 30% of my work time; three other psychologists serving similar catchment areas in the same authority reported a greater proportion of time spent on the processes contributing to statutory assessment (personal communication). For seven main grade psychologists in this service of twenty two, the statutory assessment process was the single largest element of their work. Five of the seven were the more recently appointed psychologists, so that in this service those psychologists with least experience did most statutory work (Webster et al 2000), and in providing most statutory advice had to consider most often 'how to improve problem solving of complex human issues' (DECP, 1997, Training Committee)

The importance of psychological advice is generally acknowledged in that it informs the Education Officer in decisions about provision for students, and may be referred to in negotiation about that provision (DfEE, Education Act, 1996, DfES, 2001). Negotiations arising from statutory advice may lead to tribunals, and statutory advice may be used in tribunals. Additionally, a target in producing advice in the given time scale was the only
statistical target in the Education Development Plan for the Psychological Service in this authority (Education Development Plan, X Local Education Authority, 2002). This is in keeping with the recommendations of the Association of Educational Psychologists (AEP) that a performance indicator for every service in the country should be the return of statutory advice within the required timescale (AEP, May 2002). The additional recommendation of surveys of stakeholders, including parents, children, and the local education authority, adds a qualitative measure of performance, once again relating in part to statutory advice.

Other factors may have affected my communication skills in statutory assessment besides the doctorate, particularly factors relating to central and local government objectives, for example, a consultation on the future of Educational Psychology (DfES, 2001), and a report by the Audit Commission on the process of Statementing and its relationship to indices of 'best value' (Audit Commission 2002), that is, the cost of the statement and the benefit received from it. Besides these governmental variables, it is likely that trends within the educational psychology profession have influenced my communication; the thrust towards evidence based practice will be considered (for example, Evans and Benfield, 2000: Stoiber and Waas, 2002: Frederickson, 2002).

One way of measuring my interpersonal skills has been through appraisal using 360 degree feedback. This was carried out at the start of the doctorate, and again at the finish, providing information from peers and head teachers on my communication skills, among other areas of work.

The following section, Section 2, reviews critically selected relevant literature. In Section 3 the results of the two 360 degree appraisals will be discussed, in particular those parts
relevant to communication skills and statutory assessment. It will be argued that the validity of this questionnaire in measuring communication skills in the statutory assessment process depends on a number of factors, and may be improved for future use; appraisal may also benefit from a qualitative approach to collecting and analysing data. In the final section the issues raised in the literature review, and by the appraisals, will be considered in relation to my professional development.
2.0 Literature review

2.1 Selection of literature

A search of PsychInfo, ERIC and SOSIG databases revealed little about the communication skills in statutory advice; the majority of the journal articles focused on assessment, or on the process of statementing. It was necessary to look through journals for educational psychologists, such as *Educational Psychology In Practice* and, *Educational and Child Psychology*. Only one paper referred directly to the communication skills used in statutory assessment and that was not the main thrust of the article (Buck, 1998).

However, in these professional journals there was consideration of the difficulties facing new entrants to the profession (Webster et al, 2000) relating to statutory advice, and in the context of professional development, there were papers from those journals, discussing both appraisal initiatives within their authorities and the relationship of appraisal to the job of educational psychology. It has been helpful to consider some of the earlier competencies discussed by Stratford (1994) in relation to the 360 degree appraisal used here (Sharp et al 2000), both in this section and the following section.

The recent focus on evidence based practice appears directly relevant to the content of statutory advice (see Evans and Benfield 2000, Frederickson 2002, and Stoiber and Waas, 2002) and the practicalities of this will be considered. Space does not permit the discussion of Government publications mentioned above nor of local government
priorities (Education Development Plan 2002 –2007) although their effects are acknowledged.

In Section 4 the framework provided by Webster and Hoyle using the dichotomy between the managerial demands on psychologists and the professional demands provides a useful framework to consider my professional development (Webster and Hoyle, 2000).

2.2 Statutory assessment and educational psychologists

It is argued that a paper central to this discussion was that of Webster and his colleagues (Webster et al, 2000) and it will be described in some detail. The authors considered the confidence of new psychologists who had been asked to respond to a survey on incidents they felt they had successfully handled, and those they felt they had been unsuccessful with. Ten per cent of the responses relating to success were what Webster et al referred to 'procedural' and in these was included statutory assessment. In the areas where the new educational psychologists had felt less successful, issues immediately relating to statutory advice and assessment appeared in all categories: these were skill or knowledge gaps: interpersonal episodes (as an example disagreement over an assessment is used): workload issues and unreasonable expectations of senior colleagues: ethical issues such as specialist provision: LEA systems, procedures and protocol. (Webster et al, 2000).

The authors imposed a curiously arbitrary classification of seriousness on these unsuccessful areas, going from level 1, 'uncomfortable' to level 3, where professional
mistakes are made. This level was typified by mistakes over 'ethical, political, and systemic issues' and they cited nine tribunals challenging recommendations made by these new EPs. Unfortunately there seemed to be no attempt at explaining how the classification was reached of these 'unsuccessful areas' and the authors failed to follow up the citing of the tribunals with further detail to show how and why these mistakes occurred, and why they were considered as mistakes. What they did discuss at some length were possible hypotheses for what they saw as a competency shortfall — that is, instances where psychologists were displaying level 3 mistakes. They did not consider the possibility that such mistakes are seen as serious because they take place as part of a statutory process. For example, the recommendation that a child should or should not have specialist provision as part of a statutory process carries more weight and has greater implications than the same recommendation made informally. To avoid such 'level 3' mistakes the communication skills in the giving of statutory advice is one element of interpersonal effectiveness which appears to be an important issue for pupils, parents, schools, education authorities and psychologists. It is worth speculating that some of these 'serious mistakes' may have arisen partially at least because of a failure to set out supporting evidence for arguments. The push towards basing assessment and advice upon clearly set out evidence may make such mistakes fewer.

Evans and Benfield (2000) discussed such evidence based advice, using medicine as an analogy, with a push towards collecting and evaluating all relevant medical data for reviews and dissemination of medical data. Practitioners were to evaluate the evidence themselves; the authors raised practice review groups, drawing up protocols for evidence based practice, and recommended systematic reviews with clear criteria and joint reviewing to reduce bias. As an exemplar in education, Evans and Benfield discussed the reviewing of emotional and behavioural interventions in the age range 5-
11. Only primary age studies were reviewed and a literature search was undertaken using synonyms and related terms. Database searches were supplemented with hand searches of 27 journals, which the authors noted were extremely time consuming: including reviews and experimental scrutiny, the process took about 6-8 months using 2 full time researchers in Health. Besides the research time needed, the authors raised the possible limiting effects on research of adopting a rigorously controlled scientific research model:

'Many of the reports excluded from the study were from teachers or psychologists who had adopted certain strategies with a group of pupils and who were reporting in general terms that they appeared to be effective and were recommending their adoption by others. However, these reports did not include comparison groups - - -'

(Evans and Benfield, 2000, p 538)

The push to persuade practitioners in medicine and psychology to evaluate research should be looked at critically. Firstly, are GPs qualified and/ or trained to evaluate experimental evidence? What percentage have the detailed and up to date knowledge of statistical analysis to undertake evaluation of evidence? The analogous question for educational psychologists is to ask how many can weigh up and critically evaluate an intervention for challenging behaviour, or comparative reading interventions. Scrutiny of results in papers in academic journals (often part of the psychology degree course) may indicate that effects were negligible, that preselection had taken place, that subjects whose results lay too far from the mean were dropped, or that incorrect statistical analysis had been selected. Are GPs sufficiently trained to scrutinize the sophisticated trials run by drug companies? And similarly, do EPs retain the knowledge gained in those classes to scrutinize evaluations of interventions?
Secondly, do working professionals have time to allocate to professional development to enable them to ask these questions and locate the relevant research and can professional development offer the expertise to enable psychologists to do this?:

'No longer did their professional development depend on skimming a handful of journals and attending a few conferences to receive pre digested and pre organized information'

Evans and Benfield, 2000, p 531

The third question is that of experimental control; this was based on a medical model and assumed that control of extraneous variables was as possible in educational interventions as in medical drug therapies. This assignment cannot go fully into the differences between medicine and education; however, measurements may be less exact, rating scales are more usual than nominal data, variables are more difficult to control, and judgements of behaviour may be more subject to individual prejudice and attitude. To some extent Evans and Benfield acknowledged this with a mention of trade off between 'authenticity and controlled experimentation' (ibid, 539)

This balance between the desirability of ecologically sound studies – that is, research in the classroom, - and the need to control variables, is an issue; in general there appears to be a lack of control groups and properly randomised samples, but there are pragmatic tensions in trying to recreate a laboratory situation. The theme of practice based on evidence is continued in a recent paper by Frederickson (2002) in which she has discussed some of the difficulties in applying the randomized controlled trials (RCT) of the medical field, to the field of education. While the paper recognised such difficulties it then went on to discuss a small trial of a Circle of Friends intervention. The author
recognised that the sample size was small, but suggested collaboration across services and training courses which would presumably produce much larger sample sizes, although such collaboration may carry with it the complexities of ensuring similar experimental controls and assessment practices.

2.3 **Appraisal of educational psychologists and 360 degree feedback**

The appraisal of educational psychologists has been discussed at a service level in the professional educational psychology journals (for example, Imich et al, 1992; Twisleton, 1992; Thompson and Wills, 1994; Harrison 1999; Webster, 2001). Some were accounts of introduction of appraisal systems in individual services (e.g. Twisleton, 1992; Imich et al, 1992). It appears useful to pick out from some of these accounts what is meant by appraisal. Twisleton, for example, made a distinction between appraisal, a private and negotiated scheme to address professional development and evaluation, a line management system seeking to ensure that individual work and aims are related to service aims and objectives (Tвисleton, 1992). In the case of appraisal the appraisee both chose issues to address and the appraiser, who was a peer, not the line manager, of the educational psychologist. The differences between appraisal and evaluation were made explicit in this table below:
Comparison of Peer Support, Line Management Review and Monitoring and Evaluation

<table>
<thead>
<tr>
<th></th>
<th>Peer support</th>
<th>Peer supported appraisal</th>
<th>Line management supervision</th>
<th>Service monitoring and evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process owned by</td>
<td>individual</td>
<td>individual</td>
<td>service</td>
<td>service</td>
</tr>
<tr>
<td>Process driven by</td>
<td>individual</td>
<td>service</td>
<td>service</td>
<td>service</td>
</tr>
<tr>
<td>Focus of process</td>
<td>Individual's work or issue of their choosing</td>
<td>individual</td>
<td>individual</td>
<td>service</td>
</tr>
<tr>
<td>Length of cycle</td>
<td>Ad hoc, as often as needed</td>
<td>1 year or other time scale</td>
<td>1 term or other time scale</td>
<td>1 year general issues thematic issues</td>
</tr>
<tr>
<td>purpose</td>
<td>Help!</td>
<td>Personal professional and career development</td>
<td>Job performance and improvement, service effectiveness</td>
<td>Conformity to service aims, Consumers wishes, Trends.</td>
</tr>
<tr>
<td>appraiser</td>
<td>Anyone individual chooses</td>
<td>Person of appraisee’s choice</td>
<td>Line manager</td>
<td>Management and service members</td>
</tr>
<tr>
<td>outcome</td>
<td>Feeling better, doing better, neither public nor private</td>
<td>Written private statement with targets for action</td>
<td>Joint written statement open to individual and service management</td>
<td>Changes in service development plan</td>
</tr>
<tr>
<td>Service support</td>
<td>None necessarily involved</td>
<td>Input if needed for development plus access to training</td>
<td>Staff development if needed Address organisational obstacles, input to service decision making</td>
<td>Appropriate general staff development programme input to service decision making</td>
</tr>
</tbody>
</table>

(From Twisleton, 1992, p 174)
In this table we can see that appraisal was focused on the individual and driven by the individual to help the individual in personal, professional and career development. Evaluation was driven by the service and informed the service development plan. While Twisleton (1994) acknowledged that appraisal can produce feedback information which is useful for the management of the service, he has not viewed it as necessarily linked to service priorities. There may be a possible tension between the two ends of appraisal, that of individual professional development, and the aims of the service. It is possible that such tension has increased with the requiring of 'best value' from local government services, demanding that the service to the client or user is demonstrated to be value for money through the use of performance indicators (Audit Commission, 2002).

As early as 1994 Stratford pointed out the need to satisfy the user, and set out to define what competencies educational psychologists need to perform their job to ensure quality and to satisfy the customer.

'Roles are what you have, competencies are what you do, or what you need to perform that role successfully'.

(Stratford, 1994, p 21)

However, he pointed out that self appraisal was crucial to the process, appraisal of how the self performs the professional role, as well as the perceived satisfaction of the user. Stratford's discussion is both relevant to how to decide whether an educational psychologist is improving in such areas as interpersonal effectiveness, and to the relationship between educational psychologist and statutory demands. He clustered the returns from the AEP questionnaire to produce a set of five job or task requirements, and
from these derived competencies, that is, behaviour needed to do these jobs. All of the five task requirements appeared to relate to statutory assessment, and the table below shows the tasks, behaviour, and attributes needed.

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Behaviours</th>
<th>Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Counselling</td>
<td>Listening, showing patience,</td>
<td>Empathy, caring</td>
</tr>
<tr>
<td></td>
<td>helping others produce ideas</td>
<td></td>
</tr>
<tr>
<td>Advising and influencing</td>
<td>Asserting a case, negotiating,</td>
<td>Assertiveness, self confidence,</td>
</tr>
<tr>
<td></td>
<td>marshalling arguments</td>
<td>advocacy</td>
</tr>
<tr>
<td>Investigating and observing</td>
<td>Questioning skills, observation,</td>
<td>Numeracy, data rational,</td>
</tr>
<tr>
<td></td>
<td>observation skills, data gathering/ methods of</td>
<td></td>
</tr>
<tr>
<td></td>
<td>enquiry</td>
<td></td>
</tr>
<tr>
<td>Assessing and evaluating</td>
<td>Evaluating information to make a case, decision</td>
<td>Decisiveness, clear thinking</td>
</tr>
<tr>
<td></td>
<td>making</td>
<td></td>
</tr>
<tr>
<td>Writing and administration</td>
<td>Preparing written records</td>
<td>Logical approach,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>conscientiousness</td>
</tr>
</tbody>
</table>

(Taken from Stratford, 1994, p24)
His list of competencies appears to relate to the interpersonal skills used in statutory reporting and can be used to compare to the items in the 360 degree appraisal scheme used. All require communication skills. However, Stratford did not explain how he derived the competencies and attributes from the tasks, and it would be interesting to consider whether today such a list would be considered relevant or comprehensive, today, particularly with the omission of input for school staff and groups of parents. He suggested that the competencies provided a basis for appraisal but did not explain how they could be assessed objectively. Additionally, his view of professional development seems limited:

'Ensuring continuing professional development both for those who need a retread and for those who simply need to keep up to date'

(Stratford, 1994,p28)

A consideration of 360 degree feedback described in the following section, indicates a wider remit for professional development. It incorporates feedback from those who manage the worker being appraised, peers, and those who are managed by the worker. When Webster (2001) surveyed the appraisal systems of one third of the educational psychology services in England and Wales, he pointed out that 360 degree feedback was used by a small number of educational psychology services, and tended to be used by services which had their own appraisal systems rather than using the local authority scheme (Webster 2001). Six of the forty five authorities surveyed used 360 degree appraisal, but he listed other forms of appraisal such as peer feedback, and so it may be that elements of this feedback have been used for some time. In the 360 degree appraisal system used (Sharp et al, 2000) there was peer feedback, feedback from schools, and from line managers; I used this and a description of the findings follows in
the next section, section 3, where its validity as a measure of communication skills in the statutory process is considered.
3.0 Reflexivity and 360 degree appraisal

In this section the communication skills of the educational psychologist will be considered by comparing two appraisals carried out at the start and towards the end of the doctorate. The items in the 360 degree appraisal relating to communication skills will be compared to the some of the competencies derived by Stratford (1994). The nature of this assignment is that it demands reflexivity, described as: 'the turning back of the experience upon [him or herself]' (Mead, 1934, p134) and as 'a social scientific variety of self consciousness' (Delamont 1991, p 8). The educational psychologist was here both researcher, as she considered and analysed data, and one of the participants, in rating herself on 44 items. Her values and assumptions were embedded in her rating of which items she considered most important, and her view of herself was presented in her own rating of her performance. King (1994) recommended that the history, values and assumptions of the researcher should be taken into account, as well as her relationship with the participant (herself). The picture is further complicated by the fact that the self is not static (King 1994): ‘the research situation becomes anything but straight forward ‘(King, 1994)

I have attempted to use data gathered through the use of a 360 degree appraisal for educational psychologists in a reflexive manner. Such feedback has been used as part of appraisal systems in different organizations and, as it is comprised of the views of colleagues, line managers, and those the ratee manages, as well as a self rating, it may be perceived as less subjective than appraisal using only one or two sources of feedback (London and Beatty, 1993). Sharp et al (2000) set out the case for the use of
360 degree appraisal and describe its introduction into one service, but as may be the case with all appraisal systems, some general caveats should be borne in mind. Variables other than the ratee's job performance may influence ratings; for example the behaviour of raters is influenced by the age of the rater, the liking for the ratee, familiarity, and gender bias linked to stereotypical expectations of the ratee (Warr and Bourne, 1998; Lewis and Tessele, 2000). Moreover, educational organizations are likely to show a leniency bias (Brutus, Fleeney and Londo 1998) not shown by business and military organizations.

However, it is likely to be less subjective than appraisal from one or two sources, and it is more likely to reflect the views of users if these are the raters. In general, the rating of observable behaviour such as communication skills is more likely to show convergence than other less obvious traits (Furnham and Stringfield 1998). Additionally the extent to which different raters agree in their appraisal is an important variable; interrater agreement was positively related to the effectiveness of the appraisal, particularly in educational organizations (Brutus et al, 1998). Finally, for the reflective practitioner, the agreement, or lack of concordance, in ratings from self and others, provides a useful tool to consider one's own views of oneself compared to those of others.

This 360 degree appraisal had been part of a performance review, described as a 'vehicle for implementing agreed changes to the profile of service activities and for raising the profile of the service with key user groups' (Sharp, Frederickson and Laws, 2000, p 98).

As the paper discussed the profile of one educational psychology service the emphasis here may be less on the development needs of the psychologist, but on the needs of the
service and on its visibility within the larger context of the local authority, the evaluation part of Twisleton's (1994) table. It was compiled and trialled as part of appraisal by the educational psychology service (Sharp et al, 2000), designed specifically for appraisal of educational psychologists. It is a 44 item questionnaire using two rating scales for each of the items and items relevant to interpersonal communication skills in reporting have been selected for Appendix 1. Colleagues, such as heads of schools, special needs coordinators, line managers, and other educational psychologists rate the EP on two scales for each item. One scale with 6 rating points rates the performance of the EP and the other scale with 4 points rates the importance of that item to the rater. For example, item 7 relates to the writing of reports setting out parental views.

**Importance**

<table>
<thead>
<tr>
<th>Importance</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>checks that they have understood accurately what has been said</td>
<td></td>
</tr>
</tbody>
</table>

| 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 5 | 6 |

My mean rating from colleagues was 4.9 indicating that they thought it was good, and their rating for importance was 3.4. I (self) gave myself 5.0 and thought it more important than colleagues, giving it the top rating of 4. On the other hand, a rater may feel that an item is less important and the importance rating scale is less than the performance scale. An example is item 24:

**Importance**

<table>
<thead>
<tr>
<th>Importance</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>is innovative and creative</td>
<td></td>
</tr>
</tbody>
</table>

| 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |

| colleagues | 2.3 | colleagues | 4.3 |
| self | 1.0 | self | 2.0 |
It can be seen that neither self nor others viewed this as very important; self rated the performance noticeably lower than others.

3.1 Interpersonal communication items in the 360 degree appraisal

The 44 items are divided into separate sections, one of which is called interpersonal and communication skills. My mean performance for this section was 4.8 from colleagues and 4.0 from self. (Overall ratings from self were lower than from others). Colleagues rated this section as less important than I did giving a mean rating of 3.4 while I gave a rating of 3.6. Items in the other sections are also relevant to communication and report writing and these are given in appendix 1. In both the feedbacks from my own ratings there were areas 'to fix' and areas 'to celebrate'. but none identified by others as falling within the 'to fix' category: Items in bold occur at both periods.

Earlier feedback

Areas to celebrate in 1999 from colleagues were:

- develops effective working relationships
- listens well and responds appropriately
- takes others' concerns seriously
- has good knowledge of the range and organisation of special educational support and provision
- treats people with respect
Later feedback

In 2002 the top five areas to celebrate from colleagues were:

- shows commitment to equal opportunities
- treats people with respect
- is clear about what elements of conversation and work are confidential
- always checks that there is parental consent before s/he will discuss or work with individual children
- deals with difficult situations calmly and constructively

Earlier feedback

Areas identified by self to fix in 1999 (that is the lowest ratings by self on performance) were:

- make sure s/he understands the context within which others work
- deals with difficult situations calmly and constructively
- de-escalates conflict well
- adapts his/her behaviour to help others feel comfortable
- develops effective working relationships

Later feedback

In 2002 there were again no items to fix from colleagues. From self there was

- make sure s/he understands the context within which others work
- deals with difficult situations calmly and constructively
- is up to date with current psychological research
• gives sufficient notice for any change in date/time
• explains how s/he will address needs identified

and scrutinising the data more closely another item stands out from self in 2002:

• always checks there is parental consent before s/he will work with individual children.

Lack of concordance between self and others

It is interesting that the item concerning parental consent is perceived differently by raters and self. Raters see it as being high in performance and relatively less important. It appeared that head teachers and sencos viewed the issue of parental consent differently from psychologists – as being less important, but insisted on by their psychologist. There is also lack of agreement between self and raters on deals with difficult situations calmly and constructively

Raters saw this as an area to celebrate while self saw it as an area to fix. Perceptions of behaviour clearly differed. but a functional outcome may be that if self identifies something as an area of concern, it may be a productive area of professional development: the relationship between self and others ratings are described as complex in terms of effectiveness of outcomes (Atwater et al, 1998).

How did these areas relate to the development of interpersonal skills, in particular those used in the writing of statutory advice? In 1999, it could be argued that all the areas identified as high performance by colleagues appeared to relate to it, and in 2002 the item relating to confidentiality may be important in the facilitating of communication for statutory advice. Those areas identified by self in 1999 as needing development, all
appear to relate to the process of communication in the statutory process, compared with 4 of the 5 areas identified by self in 2002. However, one of those items, number 8: -explains how s/he will address the needs

perhaps came closest to what a psychologist does in writing advice, although the addressing of needs may be done by other agencies. For this item colleagues rated performance at 4.7,(mean) contrasted with self who rated it at 3.0. Both colleagues and self thought it important at 3.3 and 3.0 respectively. Perhaps this was at the heart of the choice of this assignment as it was clearly an area of unease for self, although this does not appear apparent to colleagues.

3.2 The validity of this feedback procedure as a measurement of development or change, and methodological issues

There are several difficulties with the validity of these 360 degree data. Questions about content validity arise in that the appraisal of psychological advice in the process of statutory assessment is not specifically addressed, although many items in the appraisal used may apply to the skills needed. To measure the communication skills used in report writing the scales need to reflect this, or have content validity. No single scale item focuses specifically on communication skills used in statutory assessment although 34 of the 42 items relate to skills used for communication, assessment, and report writing (Appendix 1).

Secondly, the raters most relevant and concerned in the measurement of the communication skills of the educational psychologist in the process of statutory
assessment, would be those who use the reports. Those would be the student, if possible, the parents, the Education Officer, the head teacher and special needs coordinator of the school or provision, other agencies such as medical and social services, and other psychologists. Peer rating would be particularly relevant in considering how far analysis of need and recommendations were based on evidence. In neither of the feedbacks were Education Officers, parents, or pupils’ views sought. Feedback came from a line manager, three peers, and 6 head teachers or special needs co-ordinators. Ratings from representatives of all the user groups should have been sought.

Thirdly, methodologically, there were difficulties in analysing the data. Most importantly, the 1999 appraisal, and the 2002 appraisal could not be directly compared quantitatively as the raw data was not comparable. The analysis of the raw data had been carried out by another agency to maintain objectivity, and the data provided for the two tests was not in comparable forms. For example, the transformation of the raw data from colleagues into means for each item was not done in the earlier test. In the 2002 data all item scales are available for scrutiny. In 1999 the results were plotted graphically and the mean scale for each item was not available. Neither was it possible to compare the areas to celebrate as these were the 5 top areas from colleague ratings – that is, ordinal data, rather than nominal data. The top ratings from the earlier feedback could well be lower – or higher – that those of the later feedback. This has meant it is not possible to use it to measure the possible effect of professional development during those three years.

The use of the 360 degree appraisal had some drawbacks, in this case the failure to ensure the data could be compared over time, and to include certain user groups – children, parents and education officers.
These methodological issues are resolvable by collecting data in a similar form and retaining raw data but in an anonymised form, by using nominal rather than ordinal data, and by the selection of additional raters, particularly those whom the advice affects, and those who use the advice. It may be worth considering the addition of a section more specifically related to statutory advice.

The system is less subjective than appraisal by a line manager, may highlight discrepancies between self/other ratings to provide professional development goals, and could reflect the views of user groups. It could be combined with qualitative data (an example is given in Appendix 2): it may be that feedback from education officers, pupils and parents could be coded in the way suggested by much the same way as Pidgeon and Henwood (1996) and may yield a rich source of data.

3.3 360 degree appraisal and professional development of communication skills

How could the 360 degree feedback be used to apply more directly to those skills pertaining to statutory advice, bearing on the issues discussed in the Literature Review? Could it be used to help new educational psychologists avoid the mistakes arising in the survey carried out by Webster et al (2000)? And what items in the 360 degree appraisal relate to evidence based assessment? At this point it is useful to return to the competencies identified eight years ago by Stratford (1994). Which items in the 360 degree appraisal relate to the tasks he identified from his survey of educational psychologists?
### Table:

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Behaviours</th>
<th>Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Counselling</td>
<td>Listening, showing patience, helping others produce ideas</td>
<td>Empathy, caring</td>
</tr>
<tr>
<td>Advising and influencing</td>
<td>Asserting a case, negotiating, marshalling arguments</td>
<td>Assertiveness, self confidence, advocacy</td>
</tr>
<tr>
<td>Investigating and observing</td>
<td>Questioning skills, observation skills, data gathering/ methods of enquiry</td>
<td>Numeracy, data rational,</td>
</tr>
<tr>
<td>Assessing and evaluating</td>
<td>Evaluating information to make a case, decision making</td>
<td>Decisiveness, clear thinking</td>
</tr>
<tr>
<td>Writing and administration</td>
<td>Preparing written records</td>
<td>Logical approach, conscientiousness</td>
</tr>
</tbody>
</table>

(Taken from Stratford, 1994, p24)

Those competencies associated with statutory assessment were assessing and evaluating; that is evaluating information to make a case, and decision making. The performance criteria associated with this were:

*Evaluating information to make recommendations*

*Setting out clear arguments for and against an action*

*Summarising salient points succinctly*

Certain areas of the competencies map on to the items in the 360 degree appraisal. For example the Counselling element maps on to the Interpersonal skills, items 6-13. The Advising and Influencing tasks map on to items 15 – 27 although there is no 360 degree item which raises Advocacy, nor Assertiveness. The Investigating and Observing tasks map on to items 15 – 27, but there is nothing on Numeracy/ Data rational in the 360
degree appraisal items. The Assessing and Evaluating tasks map on to items 15-27 but the 360 degree appraisal does not mention Clear Thinking or Decisiveness. The last task, Writing and Administration, maps on to only item 38, concerned with honouring commitments, but there is nothing specific in the 360 degree appraisal on clear communication, logical approach, conscientiousness. Likewise there are areas covered by the 360 degree appraisal not raised by the competencies identified by Stratford. He has not mentioned ethics, nor using evidence (apart from numeracy/data rational), nor awareness of limits, nor knowledge of provision. Perhaps it may be conjectured that the earlier approach emphasised the need for advocacy, while the more recent scheme draws attention to awareness of limits in relation to local provision. Is this a changing focus on the role of the educational psychologist, from advocate to pragmatist, and is this related in any way to the dichotomy between managerial demands, and professional development, discussed in the next section, using the framework of professional and managerial roles?
4.0 A framework to integrate professional and managerial development

It was not possible to use the two appraisals to measure professional development during those three years, but it did reveal areas of unease for self in this area of interpersonal effectiveness - communication skills in statutory assessment. For my professional development in this area, Webster and Hoyle (2000) offered a useful framework in their discussion of the 'new professionalism', making the distinction between output measures, typically for the LEA or state, and often part of the remit of management, and outcome measures, which related to the immediate client - the pupil, or school. Managerialist approaches relating to output have been criticised by Rowland (2001):

'managerialist approaches are antithetical to sound psychological practices as they marginalise people and what is needed is a form of leadership which promotes active involvement of all people increasing the likelihood of high levels of commitment leading to ownership, empowerment and a sense of personal and professional well being'

(Rowland DECP, 100,2001 p28)

The dichotomy between managerialism and professionalism can be applied to interpersonal effectiveness in the statutory process. Evaluation of output is done by and for the service and authority, and is fairly easily quantified, – for example, the number of assessments completed within the statutory time scale, the amount of time spent in schools, the time spent in focusing on achievement in schools in special measures. As was pointed out by Sharp et al (2000) evaluation of outcome is a more complex concept. The discussion of educational psychology without the bounds of the LEA has refocused attention on professionalism, and brought about consideration of outcomes for the client.
While Webster and Hoyle recognised the increasing demands of managerialism and output measures of appraisal, they argued that professionalism would continue to be shown by a commitment to the immediate client. Such a distinction could be applied to interpersonal effectiveness skills in the process of statutory assessment. A professional, client relevant measure, is the quality of the advice, the use of evidence based practice to underlie any analysis of need or recommendations, its acknowledgement of the views and rights of child and carers, and its clarity and relevance. The providing of statutory advice is one area where the educational psychologist provides a service to the child, the carer, and the education authority, and it is argued that communication skills are central to this process.

The areas of my unease relating to communication skills and statutory assessment, may be alleviated by commitment to the rigour of evidence based advice, and to monitoring outcomes for statemented students. An example of professionalism would be a focus on the evidence based advice discussed in section 2. While there are difficulties in practitioner evaluation of evidence, it offers a return to professional use of skills, and a greater emphasis on psychological evidence and likely outcomes. Educational psychologists could have a skilled and critical role in the weighing up and giving of evidence based advice. Such an approach may offer a solution to the ambivalence with which the profession appears to regard the duty of statutory assessment.

'The future for educational psychology is not in the narrow sphere of statutory assessment of special educational needs'

(Turner (2001), Division of Educational and Child Psychology Newsletter 99 – September 2001 p9)
This theme is repeated by such comments, again in the DECP:

: 'the meal ticket we've had or statutory work is not a benefit but a hindrance'
(Wright, 1997, DECP, p64)

Lunt in the same year pointed out the difficulties for psychologists as gatekeepers to scarce resources (Lunt, 1997). On the other hand, the 1997 issue of *Educational Psychology in Practice* on statutory assessment focused on the need for excellence in statutory advice:

*without high quality consistent psychological advice and assessment must be of fundamental concern to the profession -- -- -- that such judgements and professional opinions need to be justified goes saying'*
Daly, 1997 221

To counter such comments part of my professional development in considering the skills used in statutory processes may be in considering the evidence base of advice, and the longer term outcomes of intervention for the child. *Evaluation is the means by which we ensure that our professional advice for which the House of Lords is sure we are personally responsible, enables successful outcomes to be achieved for children. - - - - Evaluation enables us to learn from our practice, as well as ensuring that we learn how effective we are being.* (Frederickson, 2001, DECP, p7)
References

Association of Educational Psychologists (AEP) (2000) SEN code of practice on the identification and assessment of pupils with special educational needs: comments from the AEP. Durham: AEP


Daly B., (1997) Psychology and LEA decision making about formal assessment Educational Psychology in Practice, vol 12, no 4 219-221


Dessent T., (1999) Educational Psychologists and their role in school improvement School of Education Exeter, School effectiveness and improvement

DETR Department of the Environment, Transport and the Regions(1999) Best Value Performance Indicators Index, Wetherby:DETR


HMI Audit Unit (1996) How good is our school? Edinburgh: Scottish Office Education and Industry Department

Imich, A., Marshall, S., Wright, C., (1992) Trick or treat? An appraisal of an appraisal. Educational Psychology in Practice, 8(2) p 107-113


164

Oxfordshire Educational Psychology Service (2001) Service to Mainstream schools 2001-2002 Oxfordshire County Council


Stratford R A competency approach to educational psychology practice: the implications for quality Educational and Child Psychology Vol.11 (3) p21-28


Webster, A Beveridge, M., (1997) The Role of Educational Psychologists in Educational Research: some implications for professional training. Educational Psychology in Practice, 13,(3) 155-165


Wright, A (1997) The move to three year training: opportunities, threats and challenges in Professional Training in Educational Psychology: the Next Fifty Years pp54-64 Frederickson, N., Lunt., I eds Division of Educational and Child Psychology 1997 BPS Leicester

165
Appendix 1

Sections in the 360 appraisal which include skills relevant to the writing of psychological advice for statutory assessment: (taken from Sharp et al. 2000)

Interpersonal and communication skills

6 responds sensitively to the needs of others
7 checks that they have understood accurately what has been said
8 explains how s/he will address the needs identified
9 does not jump to conclusions or make assumptions
11 listens well and responds appropriately
12 is honest and open in his/her communication
13 is easy to understand

Professional knowledge and practice

15 practice is based on appropriate knowledge of theory and research
16 understands theories of assessment / intervention frameworks
17 understands the strengths/weaknesses of a wide range of approaches to assessment including evidence of their effectiveness
18 is up to date with current psychological research
19 demonstrates good understanding of how to improve learning and development
20 has good knowledge of the range and organisation of special educational support and provision for children locally
21 helps people to identify and make links between all the factors which contribute to the problem situation
22 works with people to identify possible solutions to their concerns
23 works with individuals and groups to identify practical ways of addressing their concerns
24 is innovative and creative
25 makes relevant and helpful suggestions
26 shows awareness of the limits of what can be done
27 takes time to discuss how any solution will work within a particular context

Work organization and management

29 is prepared
34 checks out that others have understood what s/he means
36 is clear about who will do what, by when, and what the expected outcomes are
38 follows through on commitments
39 keeps people informed

Ethics and equal opportunities

40 takes others views into account before acting
41 treats others with respect
42 shows commitment to equal opportunities
Appendix 2

I asked the education officer I work with the following questions:

What would you like from statutory advice from me?

'Well — um — something I or J (assistant education officer) can cut and paste from. No not just - goodness - yes - that's an interesting question— we're looking at a new format for statements — something that would fit in. But you know — sometimes , there's different cases — I want you to say special provision and then others I really don't — especially if I know there's none. I think — I'm really ambivalent. I need to think longer.'

What do you think about the evidence in the report?

'Oh, that's easy — I think you're my expert — you know about that — not me. If you say something and that's the evidence that's fine — I trust what you say. But sometimes I know people use different tests — like did you hear about school saying one thing and C (an EP colleague) used a different test — so she got a different reading age.'