CHAPTER 13: SUMMARY AND SUGGESTIONS FOR FUTURE RESEARCH

13.1 INTRODUCTION

In this chapter, I shall relate a summary of my findings; I believe that the findings based on the test data are reliable as the sample size was very large but other findings which were based upon either a relatively small sample or, for other reasons, I consider less certain and I trust other researchers will explore these issues in the future. I shall conclude the work with various suggestions to any researchers who would like to engage in this fascinating field including how I would alter some aspects of my efforts over the past eight years.

I have discovered that this area of research is an extremely complex one and one which does not lend itself easily to investigation. The complexity arises not only from the variety of estimation tasks and the criteria which it is possible to use to assess it, but also from the numerous factors which may influence the estimator's ability to satisfy a given criterion. These factors include the rather obvious ones of age/experience, mathematical attainment, spatial awareness, etc. but also the affective aspects of self-confidence and tolerance for error and most of these factors are difficult to measure. It should be noted that I reduced some of the complexity of the research by limiting a great deal of it to the context of the mathematics classroom and in all tasks requiring an estimate of a measurement attribute, eg the height, of an object, the object was either present or well known to the pupils eg a double decker bus. When the pupils had to estimate the numerosity of a collection, it was either visible (a clear plastic bag of beans) or, typically, the pupils were shown a tile and asked to estimate the number of tiles that would cover a sheet of paper. Consequently, I make no claims for the universality of my findings to other contexts or relative to other problem types.

My original intention in this research was to understand the means by which pupils perform estimation and I wanted to understand the factors which caused some pupils to be more adept at estimating. I became interested in the various aspects of estimation e.g. quantitative, computational and numerical but also the assessment of estimation. The

research progressed through a process which continually revealed issues that allowed me to proceed from one area of interest to another. There were certain disadvantages of this approach as it meant that I was often pursuing a new line of reasoning although I had not `tied-up' the first line of inquiry. I wanted to explore as many aspects of the research as I could and, overall, it enabled me to avoid the danger of exploring a narrow, and potentially fruitless hypothesis. The pilot test data indicated that pupils had difficulty estimating lengths, heights, areas and very large answers, but they were able to estimate the numerosity of small collections. The analysis of the large body of test results allowed me to distil these difficulties into two areas of interest which led to the development of the Criterion Of Reasonableness (COR) and the detailed work on Imperial vs metric measure. The interviews which followed the test programme established the findings regarding the systems of measurement and the potential of the COR. The interviews, in turn, led to an interest in the affective aspects of estimation, including the valuing of the estimation process, culminating in the research into the role of risk-taking in the curriculum and the subsequent survey of the attitudes of pupils to risk-taking. In retrospect, had I limited my research to a much narrower aspect of estimation I might have made more progress in gaining a deeper and clearer understanding of that one aspect but my approach gave me the advantage of a broad view of the topic. It is from this vantage point that I intend to relate in the summary section.

It is probably a common part of the research process to realise that one might have been more productive had certain decisions been made earlier in the work. I shall make suggestions to future researchers in the final section in that spirit.

13.2 SUMMARY

I had worked with low attainers in a Research Project as part of my MSc and had been successful in improving their attainment in tasks involving estimations. I was interested in the following questions as a result of this work:

what tasks could the general school population achieve?
what tasks were difficult? and
what caused the difficulty?

The literature includes many exhortations to teach estimation and the literature search revealed a useful classification of estimation tasks, primarily through the work of Bright, and attempts to identify the processes used in estimation tasks with important contributions from Reys et al, Hildreth, Levine and Siegel et al. A major problem was discovered in the lack of a suitable standardised criterion against which estimates were assessed. Further, very few of the reports included data, which prevented me from attempting a standardisation exercise. The variety of criteria made comparison across studies difficult as most researchers had used different percentages in their criteria. This problem eventually led me to develop the COR which I shall describe later in this section.

I conducted a series of tests in the London Borough of Sutton and the analysis of this large set of test results convinced me that most pupils can satisfy the ±20% criterion to estimate the numerosity of a collection when the collection is less than 100. It also showed that many pupils have the ability to estimate lengths, areas, and volumes, within the ±50% criterion, when the magnitude of the answers to these questions is small but that when the answers become larger, the pupils fail to satisfy that criterion. Table 13A, below, shows the percentage of the school samples satisfying the ±20% and ±50% criteria for some typical questions. Questions P8(S12) & P16(S22) can be classified as numerosity problems with small magnitude answers and it is clear that a majority of the pupils can satisfy the criteria. Questions P19(S25) & P20(S26) show that metric quantities can be estimated by many pupils but I would remind the reader that the distributions for these questions showed a large number of pupils vastly overestimating some metric quantities. Questions P23(S30) & P27(S36) suggests that area and volume problems cause more difficulties but many pupils are able to satisfy the ±50% criterion. Finally, P22(S29) shows that pupils have great difficulties with a large magnitude area problem.

TABLE 13A
Percentage of Pupils Satisfying Criteria for Typical Questions

Primary	±20%	±50%	Secondary	±20%	±50%	Question Description
P8	65%	93%	S12	71%	99%	Crosses in an array .40
P16	59%	88%	S22	62%	92%	Dots in a pattern .50
P19	47%	71%	S25	42%	76%	Height of room .3 m.
P20	26%	41%	S26	29%	53%	Height of bus .4.5 m.
P23	30%	74%	S30	29%	72%	No. of blocks in box .60
P27	19%	58%	S36	24%	77%	No. of sheets on wall .90
P22	3%	17%	S29	14%	38%	No. of tiles on wall .2000

Other researchers may wish to apply other criteria to the data and I have included the bar graphs for all of the test questions in Appendix IV for that purpose. I believe that I have produced strong data showing the types of questions for which pupils **can** satisfy a reasonable criterion and these should be useful as a base from which other researchers can work. The testing programme also proved successful inasmuch as the pupils responded to the activity in a very positive manner.

The first major issue that arose from the analysis concerned the criteria for assessing estimations of large magnitudes. I consider the percentage criteria to be unsuitable for answers of large magnitudes because an estimate which, in my view, would be acceptable fails to satisfy the criterion. I have explained my reasons in Chapter 8. I developed a Criterion Of Reasonableness (COR) in response to the need for a more

useful criterion for answers of large magnitude. The COR employs existing research in perception linking the estimating of large magnitudes to the subitising of numbers less than 5 and it provides a coherent means of assessing estimation across a wide range of magnitudes. The validity of the COR has been substantiated across a wide variety of groups in terms of age and mathematical attainment. I have shown its usefulness in determining the reasonableness of an answer for quantitative estimation tasks and I consider the COR to be an important outcome of this research with a full development presented in Chapter 8.

With regard to computational estimation (as opposed to quantitative estimation), I claim that the COR offers an initial assessment of pupils' attainment but I recognise this to be a minimal criterion and other, more rigorous criteria are probably more appropriate for computational estimation tasks. I consider it unlikely that any algorithmic criterion can be developed which is suitable for all computational estimation tasks as the limits of answers which are `acceptable' vary tremendously as outlined in Chapter 9. The COR does, however, allow a potential first level of criterion which, if failed, strongly indicates the estimator needs considerable attention.

During the interviews, I was able to observe the pupils engage in quantitative estimation tasks. Many of these tasks would have been simplified had the pupils employed computational estimation techniques. However, I did not find pupils frequently using these techniques. I believe a major reason for this is that most pupils are not aware of these techniques and if they are, they do not value them. The links between computational and quantitative estimation would be an interesting study once these separate areas are better understood.

The test analysis showed that a large number of pupils vastly overestimated metric quantities and I suspected one reason for this was the absence of use of the metric system in the everyday life of the pupil. I conducted pilot interviews to develop questions which would prompt pupils to relate aspects of their lives where they used estimates of measurements outside school and from the main interview data I was able to establish that many of the pupils did not use the metric system in their lives outside school. It was clear that they were not developing their

quantitative estimation skills in that system. I have no reason to suspect the sample of pupils I interviewed to be atypical of pupils in the U.K. and teachers who have been interested in my work have written to me stating similar findings in their own work.

This would indicate that there is an urgent need to encourage pupils to develop their estimation skills in the metric system but one must also recognise the value of being cognisant of both systems. I have suggested that schools might meet the needs of the pupils by having `Metric Days' and `Imperial Days' when the pupils would work exclusively in those systems with a variety of estimation tasks.

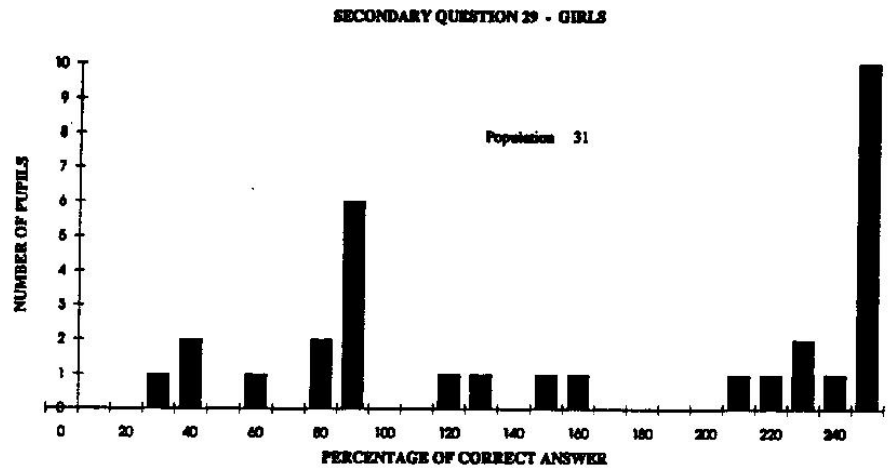
The secondary schools in the London Borough of Sutton which took part in the testing programme were either single sex grammar or single sex secondary modern schools and I had at least one of each sex of each type. Consequently, the data gave me some information on the effect of gender on the answers by pupils. I had produced bar graphs showing the distribution of the answers of pupils and I became interested in these distributions.

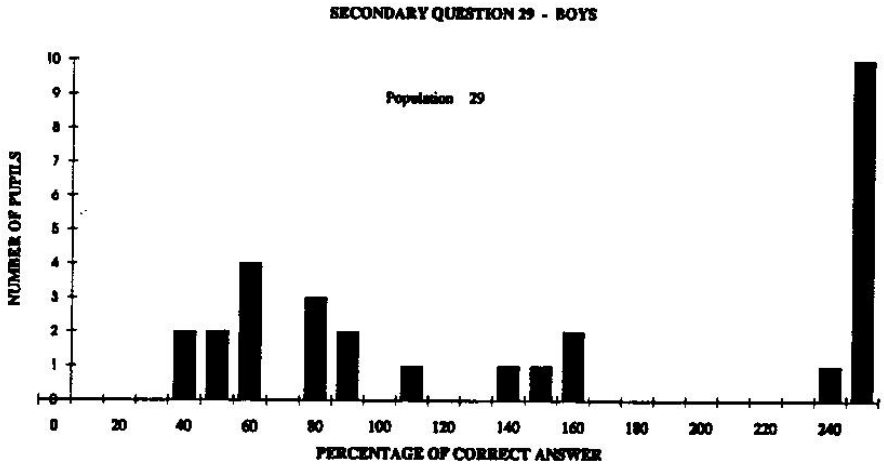
Many of the questions which produced Normal distributions for the total populations also produced many class graphs which were almost exact replicas of these Normal distributions and the success rates were virtually identical indicating that gender might not have a strong influence on the results for those questions. I have not included examples of these graphs as they would not be informative. The bar graphs (in the same format as those in Appendix IV) for a typical girls' class and a corresponding boys' class for Secondary Question 29 (How many tiles will cover the wall?) will be used to show that responses were similar for the girls and boys for a question which did **not** give Normal distributions. The reader is reminded that this question was one which many pupils found difficult. The bar graphs are shown in Figure 13.1, below and it can be seen that, though they are different, an aspect of the graphs which remains constant is the large number (approximately, one-third) who overestimated in excess of 240%. It should be noted that the percentage of pupils who satisfied the ±50% criterion were very nearly identical - 39% of the girls and 42% of the boys. The girls were more successful than the boys for more stringent percentage criteria and these differences might be worth

further investigation but, at the time, I was impressed by the number of girls/boys giving extreme overestimates which was the interesting characteristic of this graph for the total population.

FIGURE 13.1

Bar graphs for Individual Classes of S29





The characteristics which were the most interesting aspects of each of the total populations were replicated in the individual class graphs regardless

of gender. The similarities between girl and boy group graphs for each question were striking and I did not investigate possible differences in the graphs at the time. In retrospect, these differences, albeit small, might be worth exploring.

I have noticed during the years I have been interested in estimation that many pupils do not value the estimation process. I have often seen pupils estimate **after** they have made a calculation or measurement and teacher colleagues have confirmed this many times according to their experience. The importance of pupils valuing their estimate cannot be stressed enough and is one of the major deficiencies of the work set in most textbooks and schemes of work. All too often, the pupil is required to estimate and immediately measure or calculate thereby devaluing their estimate. I consider this to be one of the main impediments to the improvement in the estimation abilities of pupils in schools. An important result of my study has been the development of a method to encourage pupils to value their estimates which proved quite successful for the small sample I interviewed and a similar programme could be a useful addition to the teaching and valuing of estimation in schools.

During the course of these interviews, some pupils displayed greater confidence than others in their manner and I became interested in the level of confidence that children had in their ability to estimate. There was an indication that gender was a factor in the confidence expressed by the pupils. I developed a means by which the pupils were given the opportunity to express the level of confidence which they had in their estimate and the findings showed that, in this small sample, the girls' confidence level was substantially lower than that of the boys but their `accuracy scores' were only marginally lower than the boys. These facts may indicate that confidence and ability to estimate are not correlated, but I have difficulty accepting this idea as it appears contrary to my own classroom experience. I believe that estimation involves risk-taking and the attitude to taking a risk might be one factor in the girls' lack of confidence as societal pressures dictate that girls do not put themselves at risk whereas, in general, boys are encouraged to be more adventurous. Other possibilities are worth consideration and I shall discuss these in the suggestions for future research.

I found evidence that risk-taking is also encouraged in other curriculum areas and this prompted me to investigate the attitudes which pupils have to risk-taking. An attitudinal survey of 1320 secondary and 400 primary pupils was conducted to ascertain the effect of gender upon pupils' willingness to engage in risk-taking activities and the treatment of the data indicates that gender is a powerful factor in the attitudes which pupils displayed in their responses to the statements. Caution should be exercised in interpreting these responses as they may indicate that girls are more willing than the boys to be honest about their attitudes to the statements. The girls' greater embarrassment (according to the survey results) if they got an answer wrong may indicate that they get the answer right more often than the boys. If that is the case, then my hypothesis that confidence is linked to attainment may be incorrect.

It should be clear that I encountered more issues the deeper I got into the topic. How can one state that an estimate is reasonable? Under what criteria? How can the researcher ensure the subject is actually attending to the problem? For example, why did the mathematics educationalists at the PME conference fail to satisfy any of the criteria for the computational estimation tasks? When a pupil can only describe their estimate as a **guess** - allegedly without a valid mathematical strategy - how does the researcher gain access to the method? Do some individuals have an intuitive `feel' for the numerosity of a collection associated with general spatial awareness? The answers to some of these questions continue to elude me.

13.3 SUGGESTIONS FOR FUTURE RESEARCH

13a Introduction

Future researchers should consider the advantages and disadvantages of the various means of gathering data and I shall describe some of the problems which I encountered in this study. I utilised a large testing programme, a set of interviews and another large survey to gather data. First, I gathered a large amount of quantitative data (the pupils' estimates) which allowed me to state conclusions in the previous section with some confidence due to the large sample size. On the other hand, this data did not give me the qualitative data (the techniques used or an indication of

the pupil's confidence) which might have been useful to analyze the quantitative data for those questions which proved difficult. Secondly, I gathered some qualitative data through the interviews which was helpful in substantiating my claim that the COR is useful and that the metric system is not used by pupils outside school. Although the number of interviewees was small, I think this data is reliable inasmuch as it is not in conflict with the large body of test data and is also supported by my past experience in the classroom. Issues of confidence and attainment arose in a small set of interviews and the qualitative evidence indicated that girls lack confidence in their ability to estimate but the quantitative data to support this was very limited. Finally, the survey provided an enormous amount of quantitative data on the attitudes of pupils but did not contain a measure of attainment/performance would have been useful to provide the possibility of finding a correlation between confidence and attainment. Other researchers interested in proceeding with work on these issues will need to consider the means by which they can overcome some of these problems. A different approach to the research might be to implement longitudinal studies designed to improve the affective attributes and to link these with an intensive estimation teaching programme. This technique would allow the researcher to explore the effects of the intervention or teaching which the pupil is experiencing. The MSc Research Project was conducted more along these lines and it proved quite successful in improving the estimating abilities of the pupils involved.

One problem which the researcher faces when developing a set of questions for either interviews or tests is to ensure that they will interest the pupils and provide the researcher with useful information. Suggestions will follow in the next section as to how I would address specific shortcomings of the present study. In the interview process, a major problem, which the researcher encounters in understanding the estimation process as used by pupils, is the development of the means by which s/he can gain access to the thinking of the pupil and without intimidating the pupil. I think this can be facilitated through long term studies or teaching programmes and I shall make further suggestions in this regard in the final section.

13.3b Specific Suggestions

I would encourage the use of open-ended questions in future test programmes as they tend to give greater information to the researcher for the reasons given in Chapter 7. It is also important that the period of time allotted to the pupils to perform some computational estimation tasks is carefully monitored to prevent them using an algorithmic method; this time will probably need to be less than the 10 seconds which I allowed but the researcher must also be certain to allow sufficient time for the pupils to comprehend the problem and be able to make their `educated guess'.

I would recommend that future researchers provide access to either the raw data or provide the means by which other researchers can check the data against their own alternative criteria. The bar charts included in Appendix IV could be useful to other researchers in this regard. I regret that I was unable to enter my data into a modern database or spreadsheet but, if future researchers did this, access for other researchers would be simplified. A consensus of agreement among researchers on a criterion such as the COR might be useful to provide the beginnings of a large base of performance data.

My data showed that the pupils had difficulties with those questions relating to 2-D (area)/ 3-D (volume) which had large magnitude answers and it would be interesting to determine whether these difficulties were related to the dimensional problems or to the magnitude of the answer. To that end, I would recommend that researchers investigating this phenomenon should have some tasks which required an estimate of a single row of items (How many tiles to cover just the top row of the wall?) to enable the researcher to gain access to the pupil's ability in one dimension. It would be useful to have other questions (How many people can stand one deep along the sidelines to watch the hockey game?) which could give the researcher information about magnitude without dimensional complications. These questions could be included with the questions I used to enable an investigation of the effect of two and three dimensions. I have found that the area and volume questions were interesting to the pupils and similar questions relating to one-dimension

but with a dimensionless answer, e.g. people lining the hockey field, might be equally interesting to them.

A repeat of the primary survey could be conducted offering the three options of Agree, Disagree or Do Not Mind instead of the restricted choice of Yes or No. The reader is reminded that primary pupils were not given the neutral option of Do Not Mind. The either/or choice of Yes or No prevents the pupil who does not have a clear preference from expressing themselves.

13.3c Global Issues

Although estimation is a complex issue, I think it holds the potential to enhance an understanding of the number system because when a pupil can estimate well, s/he is well aware of the significance of place value. My data indicates the average secondary pupil's awareness of numerosity is developed at a level near 100 but I think that it is possible, due to the success I experienced with the MSc Research Project, to raise this awareness into the thousands. It may also be possible to utilise estimation to make tens and even hundreds of thousands more accessible to the majority of pupils but, even if this is not realised, there is no evidence to indicate that an estimation programme would jeopardise the pupil's understanding of number. I have found in twenty years in the classroom that many pupils identify, after units, ten, hundreds and thousands, the next column to be millions. A teaching programme to improve pupils' estimation abilities similar to that described in Appendix VII might allow pupils to gain a greater understanding of place value.

If I were to start afresh, I would concentrate on quantitative estimation tasks. I found that pupils were more willing to engage in them as they found these tasks more appealing than computational estimation tasks which are linked so strongly with numerical manipulations. I think any future researcher might find it more productive to focus their efforts on either one of these narrower aspects of estimation until a greater understanding of quantitative and computational estimation separately is gained when the relationship between the two might be a valuable line of research.

It might be useful to determine how widespread the use of the Imperial system is as a `natural' system of estimation in other countries. This could concentrate on countries which were at one time part of the British Empire but which now, officially, use the metric system. Do people in these countries use the metric system exclusively or is there a vestigial use of the Imperial system? Quantitative estimation in other countries which have used the metric system for generations could reveal other `folk' systems of estimation which might be valuable to understand the means by which people estimate measurable quantities. I suspect these `folk' systems would involve measurements that are closely linked with physical attributes of the human body. On an anecdotal level, I have often found adults in the USA and the UK who claim that the metric system does not have close links with the human body unlike the Imperial system where the length of shoe is approximately 1 ft. and the length of the first joint of the thumb is approximately an inch.

I have described `estimates' (the car body shop estimate) which are clearly not estimates in the same sense as the present research and this fact will need to be considered when selecting `good estimators'. If `good estimators' can be identified, a number of questions might be worth investigating. Are good estimators restricted to the context within which they operate as part of their work? Are good estimators of numerosity in one context (the number of people in a crowd) equally able to estimate in another context (the number of beans in a jar) and are they equally `good' at estimating lengths, area, volume and mass in a variety of contexts?

Probing the techniques used by `good estimators' might yield information about their computational estimation abilities which could prove most interesting and possibly lead to an understanding of the links between the two areas of estimation. It is possible that good estimators rely upon a vast amount of prior experience when determining their estimates. This would suggest that those individuals would not rely upon their computational skills. It is also possible that the reliance upon prior experience or computational skills varies with the individuals. I suspect that when the quantitative task is amenable to the techniques of Decomposition/Recomposition (see Chapter 4), good quantitative estimators will use computational estimation techniques. I would expect a good estimator of large numbers to operate at quite a high level of

computational estimation. The good estimator of length might be using a vast amount of prior experience to make their estimate without any of the techniques of computational estimation but then again, that could be the case with the crowd estimator. A rather interesting problem, to say the least!

The literature search revealed that several studies had been made with good estimators. I realised that estimation tasks required a variety of techniques and abilities but when I tested my questions on mathematics educators (the PME conference, BSRLM meetings, and a King's College seminar), I assumed that these subjects would be good estimators in all areas. This assumption was clearly flawed as the PME data showed. I would suggest two possible explanations for the `failure' of the PME participants to satisfy the criteria may be:

1) they did not attend to the task; they were asked to complete the task in a busy hall with many other exhibits surrounding them,

2) it is possible that some were experts in psychology and not as well versed in mathematics as I expected.

I did not take the affective domain into account in the early stages of the research but I think that efforts to ascertain how affective aspects impinge on the estimating process would be very interesting and useful. For example, I believe that self-confidence is of importance in the pupil's self-image and that this could have an effect on their ability and willingness to estimate. The evidence from my sample showed the girls' attainment was only marginally less than that of the boys but their confidence level seemed to be much lower and this appears to contradict my hypothesis. I am very interested in possible explanations for the girls stated lack of confidence and would like to suggest several possibilities in the short list that follows:

1) it is possible that the girls' stated lack of confidence may be a result of their greater willingness to admit to a lack of confidence,

2) the girls reacted to my gender in a negative manner,

3) the sample was very small and a larger sample might show different results, or

4) it may be a genuine contributing factor for their slightly lower level of attainment and with greater self-confidence the girls may be able to estimate with much greater accuracy.

If 4) above were the case, one might predict that if girls had equal confidence to the boys, they would be much better estimators. In any case, it would be useful to examine performance and attitude together in any future study of the relationship between confidence and estimating.

Further research might be interesting to determine how self-confidence affects the willingness and ability to estimate. My later interview questions obtained an element of measurement of the confidence the pupil had in their estimate and a measure of their ability, but did not establish the pupil's general level of self-confidence. In other words, some pupils may exhibit great confidence in an estimate but compared to their general self-confidence, their confidence in the estimate is quite low. It might be valuable to utilise any experience gained in other disciplines of measuring self-confidence and relating this to the pupil's confidence in estimating.

I believe that a strong factor related to the willingness to estimate is that the estimator must be aware that they cannot be exactly correct but they also must be able to tolerate a level of error and a measure of this tolerance could be valuable for future researchers. Again, the problem of gaining access to the pupil's thought processes arises. If the pupil perceives that the researcher is implying that their answer is `wrong', the pupil may resort to avoidance tactics and not provide the researcher with the required data. These problems are self-evident but they do show that the researcher must always be wary of their position in the interview process.

I have hypothesised that the previously described `estimating pupil' should be more willing to take risks in her/his mathematics with the potential that this will be beneficial in developing a more positive approach to problem solving. The survey data showed that gender is a factor in risk-taking but the interview data related to confidence and accuracy might indicate that girls are not severely disadvantaged by this unwillingness to take risks. If a means of assessing a pupil's self-confidence and tolerance for error were developed, it could be useful to an improved understanding of the affective aspects of estimation. It might also be useful to determine the relationship, if there is one, between pupils' confidence level and their attainment in a range of contexts with

particular attention to gender. This entire area appears to have a potential for some valuable research.

There exists the possibility of drawing together all of the strands of the efforts I have suggested. The criteria for assessing estimation tasks could be developed utilising the information provided by the good estimators. I envisage longitudinal studies of groups of pupils partaking in programmes of work designed to improve their tolerance of error and/or self-confidence run concurrently with the estimation teaching programme. I suggest that the effect of such a set of programmes could be to develop a greater tolerance for error within the pupil, thereby increasing their willingness to estimate, and, possibly, improving their estimating ability. One other effect that I would like to encourage is the pupil's valuing the estimation process which I think would be a natural result of the programmes. Finally, I think the development and implementation of a programme of work to give pupils the means of undertaking `risky' activities within a `safe' environment might make valuable use of the findings of the attitude survey. Careful preparation of suitable material would be required but, if such a programme could be developed, it might be a valuable addition to an improvement in pupils' general self-confidence with potential positive effects elsewhere in the school.