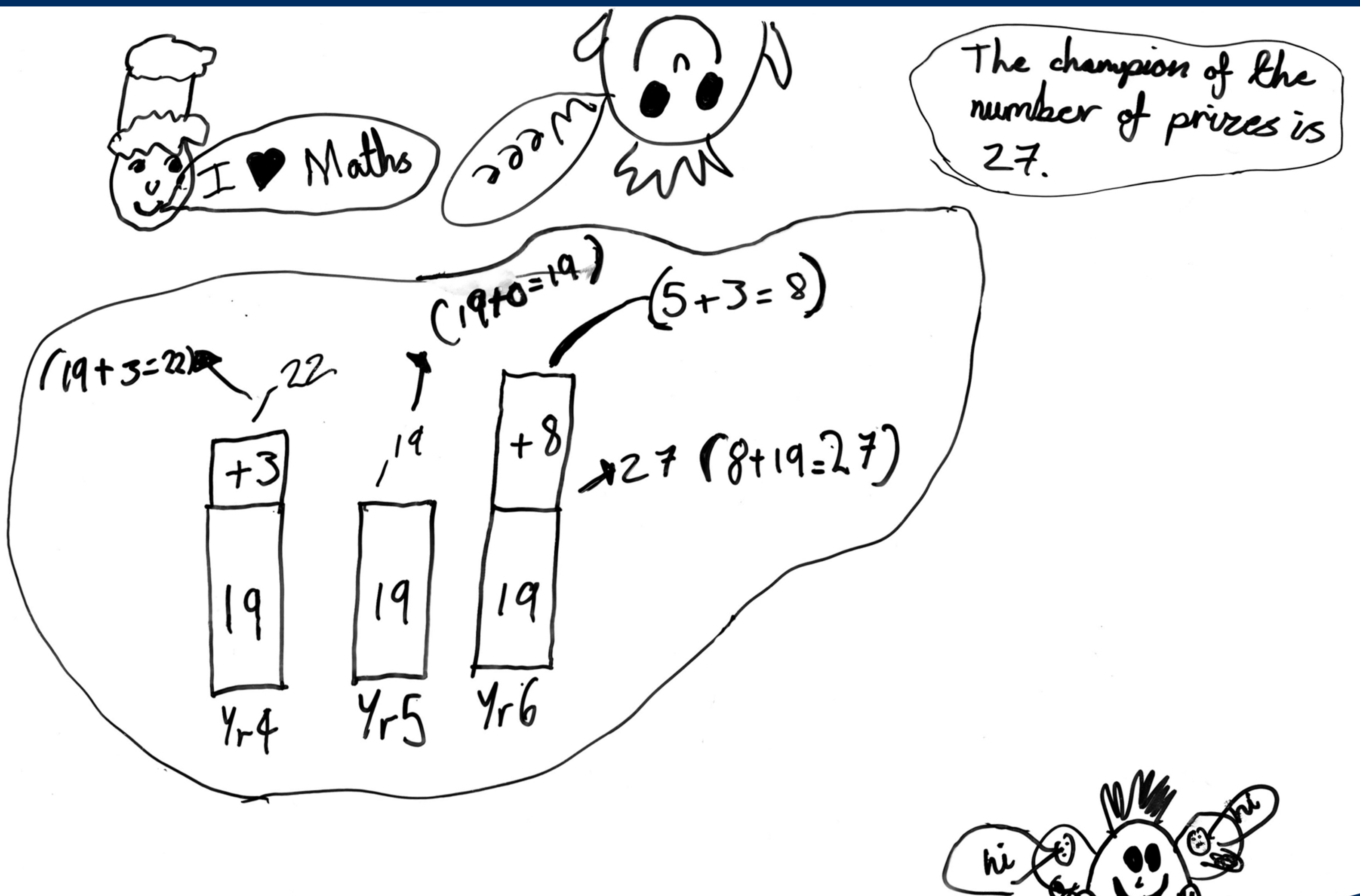
Department of Education
University of Oxford

Teaching mathematical reasoning: Probability and Problem solving in Primary School

Terezinha Nunes, Peter Bryant, Deborah Evans, Laura Gottardis and Maria-Emmanouela Terlektsi



TEACHING MATHEMATICAL REASONING

Probability and Problem Solving in the Classroom

Terezinha Nunes
Peter Bryant
Deborah Evans
Laura Gottardis
Maria-Emmanouela Terlektsi

University of Oxford

Acknowledgements
We are extremely grateful to the schools, teachers and children who participated in this project. They contributed large amounts of time and commitment to the work as well as feedback on the assessments and learning activities. We are very grateful to the Nuffield Foundation for the funding received for this project.

The Nuffield Foundation is an endowed charitable trust that aims to improve social well-being in the widest sense. It funds research and innovation in education and social policy and also works to build capacity in education, science and social science research. The Nuffield Foundation has funded this project, but the views expressed are those of the authors and not necessarily those of the Foundation. More information is available at www.nuffieldfoundation.org

The handbooks to support teachers in the implementation of the resources described in this report and the assessments can be downloaded from:
http://www.education.ox.ac.uk/research/child-learning/resources-2/#lm

TEACHING MATHEMATICAL REASONING:

PROBABILITY AND PROBLEM SOLVING IN PRIMARY SCHOOL

Contents

Executive summary 1
Key Findings and their Educational Implications 4
    Teaching Probability in Primary School 4
    Teaching Problem Solving in Primary School 4
    The Specificity of Mathematical Reasoning about Probability and Non-probability Problems 5
    The Probability and Problem Solving Assessments 5
Introduction 7
The First Phase: a Design Experiment 14
Teaching Probability 16
    The Probability Teaching Programme in the First Phase 16
    The Effectiveness of the Probability Teaching Programme 21
        Understanding of Randomness 22
        Working Out the Sample Space 25
        Quantifying Probability 25
        Summary 27
Teaching Problem Solving 28
    The Problem Solving Teaching Programme in the First Phase 28
    The Effectiveness of the Problem Solving Teaching Programme 37
        Understanding the Inverse Relation Between Operations 38
        Understanding Relations Between Quantities 40
        Summary 40
The Second Phase: Probability and Problem Solving in the Classroom 41
    Probability Results in the Second Phase 42
    Problem Solving Results in the Second Phase 45
Discussion and Conclusions About the Whole Project 48
References 51

EXECUTIVE SUMMARY

Research on mathematical learning has made it increasingly clear that children's ability to reason about mathematical problems plays an extremely important part in the progress that they make in this subject at school. To solve mathematical problems, including problems that are not used routinely in the classroom, one needs to be able to reason about the problems as well as to calculate their solutions. This is true even of apparently quite simple problems. Take, for example, the problem "Billy and Jason have the same amount of money as each other: but then Billy gives Jason £10. How much more money does Jason have than Billy?" This problem was presented to over 1,000 children aged 9 to 11 years. Only about a quarter of them solved it correctly. This is a hard problem, not because of the calculation that has to be done (10+10), which is quite easy, but because to solve the problem the child has to reason that Jason now has £10 more and Billy now has £10 less; these relations have to be added without knowledge of the quantities.

Different kinds of problems make different sets of demands on people's mathematical reasoning. One notable example of this is probability problems. The need in our society to understand probability is indisputable. This understanding is basic to scientific and statistical literacy and also for thinking clearly about risk, an important concept in science and everyday life. Consequently, a crucial question for education is whether learning about mathematical reasoning in the context of problems that do not involve probability can improve pupils' mathematical reasoning about probability.

Understanding probability involves specific concepts. To solve probability problems, people must understand something about the nature of randomness and must also realise that the probability of a particular outcome is a proportion between the frequency of that outcome and the combined frequency of all the possible outcomes in the situation being considered. Understanding probability also involves mathematical reasoning of a more general nature: it requires being systematic and following a line of reasoning to its conclusion. Thus another crucial question for education is whether learning to reason about probability problems improves mathematical reasoning more generally.

Probability used to play a small part in primary school mathematics, but it has not been included in the new primary school mathematics curriculum in England. There are at least two reasons for this omission. One is the widely held view that probability concepts are too hard for children of primary school age, but this view is probably too pessimistic. Some important aspects of reasoning about probability are a common part of the everyday lives of primary school children. An example is randomisation, which children experience in the form of shuffling cards, throwing dice and tossing coins as part of games. The other main reason for excluding the topic of probability from the new primary mathematics curriculum was because it was thought that more time should be given in the already crowded mathematics curriculum to basic mathematical topics. Yet, some aspects of probability, such as working out proportions, are part of the school mathematics curriculum anyway, and there seems no reason why they should not be taught in the context of probability as well.

There is therefore a clear case for research on the feasibility of teaching probability systematically in primary schools just as there is also an obvious case for research on teaching mathematical reasoning in the context of non-routine problems that do not involve probability. The two sorts of mathematical problems have something in common: all mathematical reasoning problems involve some preliminary reflection on the quantitative relations involved in the problem. It may well be the case that teaching children how to reason about one kind of mathematical problem will affect how well they cope with other kind.

For these reasons we designed a two-phase project in which we assessed 9 - 11 year old children's understanding of probability and their ability to reason about and solve other mathematical reasoning problems.

In the first phase of the project, the researchers taught the children either about probability (the *Probability group*) or about reasoning in non-probability problems (the *Problem Solving group*). The children were given 15 teaching sessions, spread over three school terms. All the children also took pre- and post-tests, which consisted of a series of probability and non-probability problems for them to solve. The effectiveness of the teaching programmes was evaluated by comparing the improvement from pre-test to post-test of the children in the two groups in solving the problems in these assessments with that of the children in a *Comparison group*, who attended normal lessons, but did not participate in either teaching programme. At the end of this phase, the teaching programmes were assessed and modified in the light of the children's responses during the teaching and in the pre- and post-tests.

These modified teaching programmes were then used by classroom teachers in their own classes in the second phase of the project. One set of teachers taught their children about probability and another set taught their children about solving non-probability problems. In this phase all the children were also given pre-tests before and post-tests after the teaching in order to measure the improvement in solving both kinds of problem. They were asked to try to use one mathematics lesson a weak for teaching the materials in the programme. The interval between the pre- and the post-test varied from about 6 to 8 months.

KEY FINDINGS AND THEIR EDUCATIONAL IMPLICATIONS

Teaching Probability in Primary School

- Results The results of the probability teaching programme were positive. The children who participated in this teaching in the first phase improved significantly more than other children in their understanding of randomness (for quantitative information on the results, see full report for statistical details) and of how to work out the sample space systematically. There were no significant effects of the probability programme on the children's quantification of probability scores in this phase. When the revised programme was used by teachers, it was even more effective: children in the Probability group improved significantly more than the other children in their understanding of sample space and, this time, in their proportional quantification of probability as well.
- Implication These positive results establish that it is possible for teachers to teach their pupils at the end of primary school about the basic aspects of probability systematically and successfully. We urge that this result is taken into account in future reviews of the mathematics curriculum.
- Output The programme for teaching probability is readily available to any teacher who would like to use it and the topic could easily be included in the mathematics curriculum. It can be downloaded from: http://www.education.ox.ac.uk/research/child-learning/resources-2/#lm

Teaching Problem Solving in Primary School

- Results The problem solving teaching programme was also successful. The children who participated in this programme in the first phase improved in their use of the inverse relation between addition and subtraction to solve reasoning problems more than children in the other groups. When the teachers used the modified problem solving programme in the classroom, it led to significant improvements in the understanding of the inverse relation between operations and in a general assessment of mathematical problem solving (see full report for information on the statistical analyses).

- Implication These results establish that primary school teachers can actively teach mathematical reasoning to children.
- Output The problem solving programme is also readily available and free, and the activities can easily be incorporated into regular mathematics lessons. It can be downloaded from: http://www.education.ox.ac.uk/research/child-learning/resources-2/#lm

The Specificity of Mathematical Reasoning about Probability and Non-probability Problems

- Results There was little evidence of transfer of the newly learned reasoning skills between the programmes. It is not possible to rely on children applying their growing mathematical reasoning skills to understanding probability without specific teaching of probability. Nor was there any evidence that children will spontaneously use the systematic approach learned in the context of probability problems to solve mathematical problems that do not involve probability.
- Implication A comprehensive programme for teaching mathematical reasoning in primary school should include lessons about probability and non-probability problems. The results of combining these two types of instruction have not been assessed in this project and could be of significant impact.

The Probability and Problem Solving Assessments

- Results The pre- and post-tests were assessments of the children's understanding of probability and problem solving ability. Each of these assessments showed good construct validity and reliability over time. Part of the problem solving test has been shown to be a good predictor of end of Key Stage achievement in mathematics.

- Implication The assessments can be used for helping to define learning outcomes in problem solving and probability in school.
- Output For the first time an assessment tool has been developed covering different aspects of mathematical reasoning that can be used by teachers in the last two years of primary school. The assessment is straightforward and easy for teachers to administer to groups of 30 pupils. All the resources needed are provided and freely available. The resources can be downloaded from: http://www.education.ox.ac.uk/research/child-learning/resources-2/#lm

TEACHING MATHEMATICAL REASONING:

PROBABILITY AND MATHEMATICAL PROBLEM SOLVING

INTRODUCTION

Mathematical problems can tax children's intellect in different ways. In some problems, the actual calculations that children have to make to reach the right answer are obvious and the only difficulty that they might have is in doing the calculation correctly. In other problems, it is quite hard for children to work out what kind of calculation they need to do to solve the problem, even when the calculation itself is an easy one. To solve this second kind of problem, the child has to begin by reasoning about the quantities involved and this reasoning, which we call *mathematical reasoning*, is always about the relations between different quantities or between different states of the same quantity over time.

Here is an example of what we call a mathematical reasoning problem because it can only be solved by reasoning about the quantitative relations involved in the problem before the actual calculation is done. "Billy and Jason have the same amount of money as each other: but then Billy gives Jason £10. How much more money does Jason have than Billy?" We presented this problem to over 1,000 children aged 9 to 11-years (mean age 10 years 1 month). Only about a quarter of them solved the problem correctly; two thirds answered that Jason had £10 more than Billy. This would be right if only one of the quantities had changed – if Billy had given £10 to someone else, not Jason – but in fact the story is about two changes, one in Billy's amount of money and one in Jason's. If Billy gave £10 away, he has £10 less than before, and Jason has £10 more than before. The calculation needed, 10+10, could hardly be simpler, but the solution, requires the children to deduce two relations and operate on these relations without knowing the quantities. Children's difficulties with this problem, and many other kinds of mathematical reasoning problems, led us to the conclusion that teaching them about mathematical reasoning, and showing them how important this reasoning can be, might help them through an important part of their mathematical learning.

This need for such teaching is now quite widely recognised. The new mathematics curriculum for primary schools in England proposes that “A high-quality mathematics education ...... provides a foundation for understanding the world, the ability to reason mathematically, an appreciation of the beauty and power of mathematics, and a sense of enjoyment and curiosity about the subject" (Department for Education, 2013, p.3). The importance of learning to reason mathematically is also bolstered by strong evidence for the crucial role that this learning plays in children’s mathematical achievement in school. In one of our studies (Nunes, Bryant, Barros, & Sylva, 2012), over 2,000 children in the age range 8 to 9 years took a test of mathematical reasoning that consisted of problems that required them to reason about quantitative relations and could be solved by very simple computations. Their scores in this test were then related to their performance in the Key Stage 2 and 3 mathematics assessments over the subsequent five years. We found a strong relationship between the children’s scores in the initial reasoning test and their success later on in the Key Stage mathematics assessments, even after we had used stringent statistical controls for the effects of IQ differences among the children. The connection between mathematical reasoning and achievement in the Key Stage tests was stronger than the connection between arithmetic skills and Key Stage achievement.

Where does the ability to reason mathematically come from? It is possible that children’s informal learning outside school plays a part: how well they reason when they start school may be the basis for their mathematics learning in school. There is already some evidence to support this idea: children's mathematical reasoning at the time they start school is a good predictor of how much they achieve in mathematics in school (Nunes, Bryant, Evans, Bell, Gardner, Gardner, & Carraher, 2007). However, it is unlikely that informal learning is the only source of mathematical reasoning: children may also need to learn new ways of thinking mathematically in school. Teaching mathematical reasoning aims, among other things, to help children to be systematic in their approach to problems, to pursue a line of enquiry to its logical conclusions, and to use mathematical tools that have been developed to support this pursuit in ways that are uncommon outside school.

A notable example of mathematical reasoning that needs to be taught in school is about probability. Children's, as well as adults', informal knowledge about probability is notoriously limited, largely because reasoning about probability often rests on the comprehensive construction of all possible outcomes of a situation, and this systematic approach may not be part of informal mathematical experiences. Take the well-known “two-dice problem” (Fischbein & Gazit, 1984: Fischbein, 1987): what is the probability of obtaining (for example) a total of 7 when one throws two dice at the same time? In order to approach this problem sensibly, one needs to work out systematically and to write down all the different ways in which the two dice could fall (say the first falls on 1 and so does the second; the first falls on 1 and the second on 2; and so on), to count how many different possible outcomes there are altogether and also how many times the total of the two dice in each of these combinations comes out as 7, and then to express the relation between these two numbers as a fraction or a percentage. (In fact, the answer is that 6 of the 36 possible combinations add up to a total of 7, which means that the probability of throwing that total is 6/36, or 1/6, or 17%.) The arithmetic required for the solution is rather simple; the difficulty of the problem lies in the need to be systematic in order to generate all the possible combinations and to express the relation between the favourable cases (i.e. the number of times the dice add to 7) and the total number of possible outcomes as a fraction.

Reasoning about probability is currently viewed by many (e.g. Dawes, 2001; Gigerenzer, 2002) as vital for scientific and statistical literacy as well as for dealing with a variety of situations in everyday life. Thus probability is a key concept in mathematics, but it is not included in the new English primary school curriculum. One reason for this omission is that many people think that understanding probability is beyond the grasp of primary school children because even adults often struggle with probability.

The decision to omit probability from primary school mathematics curricula cannot be based on sound research on teaching children all the basic aspects of probability since no such research study has ever been done. Yet, research on the possibility of giving children a systematic introduction to probability at this stage of their school career is badly needed for at least two reasons. One is that primary school children might very well be able to grasp some of the basic aspects of the analysis of probability if it is taught in a way that they understand. For example, most children of this age are familiar with randomisation in one form or another, like shuffling cards or throwing dice, and it seems likely (though we can find no research on this) that they will quite easily understand the relation between randomness and fairness. Other aspects of probability might prove more difficult for them: for example, a basic and entirely necessary first step in reasoning about the probability of a particular outcome is to work out what all the possible outcomes are: these make up what is called “the sample space” and figuring out this space usually requires some systematic thinking, known as combinatorial reasoning. But even if primary school children cannot master probability completely, it is still possible that they could increase their understanding of probability by systematic teaching.

A second reason for looking into the possibility of teaching children about probability is that the forms of reasoning that children have to learn in order to solve probability problems may also help them to solve mathematical reasoning problems in general. In a previous project, we established that children's performance in a mathematical reasoning test containing problems that were not about probability was highly and significantly correlated with their performance in a probability assessment. Learning to be systematic and to follow a line of reasoning to its logical conclusions in the context of probability problems could expand children's conception of what is mathematical reasoning in other contexts as well. Schoenfeld (1988) noted that a major obstacle to problem solving among adults was the belief that, if you understand the concept involved in a mathematical problem, most problems are solved in five minutes or less. However, if children were to work on probability problems that require them to generate the sample space before they can address the question, they may see that exploring the meaning of a problem is an essential part of solving it. They may learn to think things through before doing calculations even when the problem is not about probability.

The possible link between learning to reason about probability and solving mathematical problems in general led us to the idea of designing two different teaching programmes and comparing their effects. The purpose of one programme would be to teach children to reason about probability. The other programme would be about solving mathematical reasoning problems that do not involve probability. This comparison would tell us whether teaching children to reason about probability improves just their understanding of how to solve probability problems or also affects how well they can solve mathematical reasoning problems that are not about probability. The comparison would also show whether a problem solving programme that does not involve reasoning about probability improves children's solutions to probability problems or whether children need teaching about concepts specific to probability situations, such as sample space, in order to solve probability problems.

In order to design an effective problem solving teaching programme, we drew on two ideas about children's difficulties with such problems. The first relates to children's difficulties in reasoning about the inverse relation between arithmetical operations. For example, Verschaffel (1994) gave Belgian 6th graders (aged about 12) the following problem: "Pete has 29 nuts. Pete has 14 more nuts than Rita. How many nuts does Rita have?” In this problem, the comparison is stated as “14 more nuts” but the inverse relation "Rita has 14 nuts less than Pete" is the one that helps children understand that the answer is obtained by subtraction. The rate of correct responses for this problem was 71%. Results for children in England in the age range 11 to 13 show a rate of correct responses between 60% and 67% to a problem that similarly involves the inverse relation between addition and subtraction (Hart, Brown, Kerslake, Küchemann, & Ruddock, 1985).

The second idea relates to the more general point that children need to learn to think about relations between quantities and to explore these relations systematically before implementing a calculation. A method, adopted by schools in Singapore (Ng & Lee, 2009) and Japan (Murata, 2008) to support reasoning about relations between quantities, is to use bar diagrams to represent the quantities and to arrange the bars in the diagram in ways that express these relations. These diagrams are viewed as helpful for the solution of non-routine problems - i.e. problems that do not conform to routines that are familiar to the children either in the classroom or in their everyday lives. An example is: "Kate, Donna and Jamie shared some stickers between them. Altogether they bought 22 stickers. Donna got 3 more than Kate and Jamie got 4 more than Kate. How many did each of them have?" This type of problem, known as "unequal sharing", is often used as part of pre-algebra exercises and is difficult for 9- to 10-year-olds. We presented it to the same sample of children mentioned earlier on; about 31% of the children answered it correctly. This is a non-routine problem because, in routine sharing problems, the quantity that each recipient receives is the same, and the operation used is division. However, in unequal sharing problems pupils have to think about how to handle the inequality of the shares and some of them reject division as an operation relevant to this problem because the shares are unequal. Can children be taught how to reflect on non-routine problems like this one and broaden their understanding of mathematical reasoning to include the systematic analysis of relations between quantities in a problem? Will they be able to extend their learning in this context to probability problems or will the concepts specific to probability, such as sample space, remain outside their problem solving skills?

The first starting point for the Nuffield-funded project that we are describing here was a review, also funded by the Nuffield Foundation, on how children learn about probability (http://www.nuffieldfoundation.org/sites/default/files/files/NUFFIELD_FOUNDATION_CUoP_SUMMARY_REPORT.pdf). The review showed that some of the cognitive demands involved in understanding probability are shared with other types of problem that can be solved by primary school children (for an example, see Nunes, Bryant, Evans, Gottardis, & Terlektsi, 2014). This encouraged us to design a programme with the aim of promoting young children's understanding of probability.

Our second starting point was another review of research, this time on children's problem solving in general (Nunes & Bryant, 2009), which was funded by the Nuffield Foundation (http://www.nuffieldfoundation.org/sites/default/files/P4.pdf). The review provided the basis for the development of the problem solving programme and its assessment.

A research grant from the Nuffield Foundation gave us the chance to find answers to the questions we had concerning the connection between understanding probability and problem solving. With this grant, we designed and carried out a project on teaching children of 9 to 11-years to reason mathematically, either in the realm of probability or outside. In this two-year project we developed and tested two programmes for teaching mathematical reasoning, one on probability and the other on mathematical problem solving. Although both programmes aimed to improve children's mathematical reasoning, the two differed in terms of the kind of mathematical reasoning that they dealt with. One teaching programme involved reasoning about probability problems and the other focused on non-routine problems that were not about probability. The children who took part in the project participated in either one or the other of the two programmes – never both.

Both teaching programmes were entirely new. In order to evaluate them, in the first phase we taught and observed the sessions ourselves. From the lessons that we learned by doing so, we were able to adjust and improve both programmes during the first phase and after it was completed. As a result, we were able to present teachers in the second phase with two revised programmes, which they could use in their own way with their own pupils. As in the first phase, the children participated in either one or the other of the two programmes. Each of these phases is now described separately.

THE FIRST PHASE: A DESIGN EXPERIMENT

The first phase took the form of a "design experiment". In design experiments, new intervention methods are tried out, their effects are monitored, and the intervention methods are adjusted accordingly during the study (Brown, 1992; Lobato, 2003; Cobb, Confrey, diSessa, Leherer & Schauble, 2003). In this phase, we observed and recorded the children's answers and reactions to the new activities during the intervention sessions, and we monitored the success of the interventions in five detailed assessments given during the intervention period. We adjusted what was done in the intervention and also the assessments, in the light of these sources of information.

Three primary schools and a total of 73 children participated in this phase. In each school, the children were randomly allocated to one of three groups:

1. *The Probability group* worked on probability problems (24 children)
2. *The Problem Solving group* worked on non-routine mathematical reasoning problems (22 children)
3. *The Comparison group* did not take part in either teaching programme (27 children).

We worked with small groups of eight or nine children in 15 weekly teaching sessions that lasted for about 50 minutes. Our teaching programmes were run at the same time; the nine children in each school in the Comparison group remained in the classroom while the others participated in our programmes. The children were in their last term – the Summer term - of Year 5 when the project began.

This phase lasted for three school terms.

1. *Term 1* The project started with a pre-test in the Summer term (Time 1) which contained items which assessed the children's understanding of probability and their ability to solve mathematical reasoning problems that involved the inverse relation between operations. The intervention started a week after the pre-test and continued for 5 weekly sessions. A week after the end of these sessions the children were given a second assessment with similar items (Time 2).
2. *Term 2* In the following Autumn term, the children, now in Year 6, were given a similar assessment test (Time 3) before the intervention resumed for a further 5 sessions.

3. *Term 3* Before the intervention resumed for another 5 weeks in the following Spring term the children were given a further assessment (Time 4). After the intervention ended, the children were given a final assessment (Time 5) containing probability and mathematical reasoning problems.

The reason for there being as many as five assessments in the first phase was that we needed to monitor different aspects of the interventions and of the assessments as well, so that we could adjust these during the first phase in order to maximise the probability that the second phase assessments and interventions would be effective and appropriate.

The approach to teaching the two programmes was as similar as possible: we began each session by dividing the children into pairs and then giving all the pairs the same set of problems to work on together. As a motivating factor, these problems were often set in the context of games or presented by projecting computer images on a screen. After completing the problems, each pair of pupils presented its own solution to the rest of the group, either with the help of an overhead projector or on a white board. The pairs discussed their solutions with the whole group and with the researcher who was directing the group.

TEACHING PROBABILITY

The Probability Teaching Programme in the First Phase

The teaching of probability was based on our idea that three basic concepts underlie the understanding of probability. These concepts are not steps in a sequence; understanding probability involves the simultaneous coordination of all three concepts.

One is the concept of **randomness**, which is an essential ingredient of all probability problems. Random events are unpredictable, in that one never knows exactly what will happen next in a random sequence.

Another is the concept of the **sample space**. This useful term means all the possible outcomes that there are in the situation that you are considering. For example, in the *"two-dice" problem,* which we have mentioned already, the sample space is the 36 different and equiprobable ways in which the two dice could fall. This is the basic sample space for the problem, but it can be categorised (the technical term is "aggregated") in various ways, one of which is by the possible totals of the two dice. This aggregation-by-totals produces outcomes that are not equiprobable. The probability of the two dice adding up to 7 is actually 6 times the probability that they will add up to 12 or add up to 2. This is because there is only one way of throwing a total of 2 (both dice coming up as 1), but six ways of throwing a total of 7 (1 and 6, 2 and 5 and so on).

The third basic concept concerns the **quantification of probability**. This is always a proportional calculation: you must work out the proportion between all the possible outcomes and the individual outcome that you are considering, which is often given as a fraction, a proportion or a percentage. In 6 of the 36 possible outcomes in the two-dice problem, the numbers on the two dice would add up to 7: so the probability of throwing a total of 7 is 6/36 or 0.17.

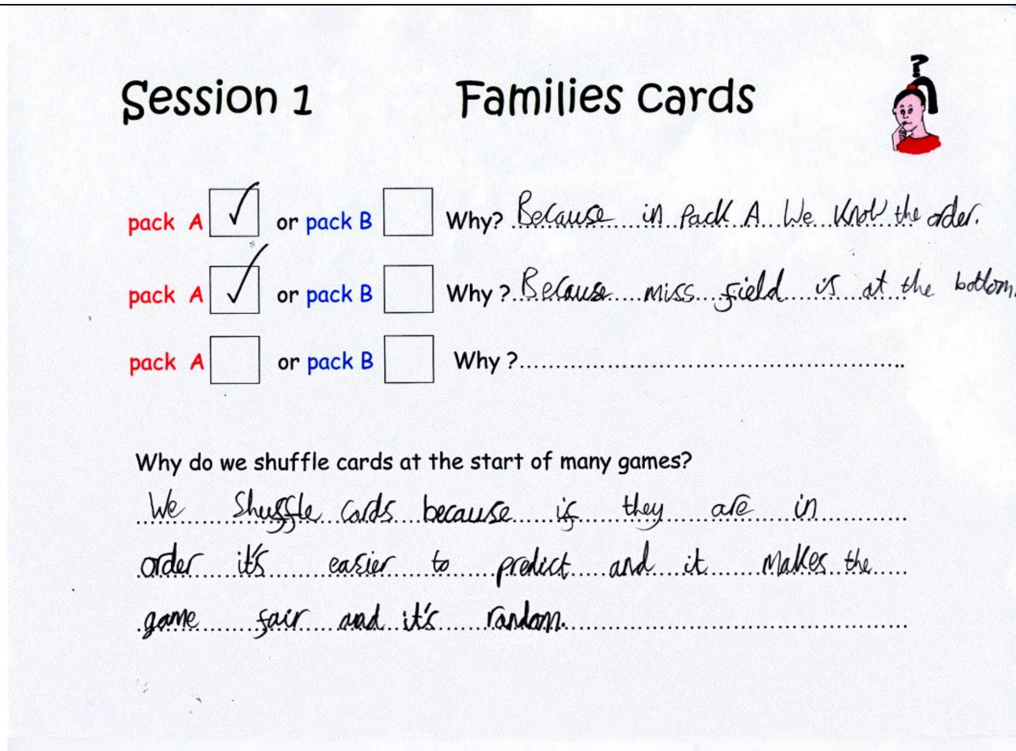
Our approach to helping children to learn about these three concepts was to dedicate five of the 15 lessons to each concept, but in some of the lessons the children were given problems that required the use of two or three of these concepts together. For example, when the children solved the two-dice problem, they had to consider at the same time both the sample space and the quantification of probability. In the sessions on quantifying probability we concentrated on expressing probabilities in ratios because previous work (Piaget & Inhelder, 1975; Nunes &

Bryant, 1996) had suggested this as the easiest way of introducing pupils to proportional reasoning.

In every session, separate pairs of children in each group worked actively on solving particular problems with materials that are relevant to their everyday experiences. Our emphasis was always on active discussion by the children themselves about what they were doing. At first the children in each pair discussed the problem with each other to arrive at a joint solution, and later on the pairs presented their solutions to the rest of the group, which led to group discussions that were often lively, even heated. We encouraged the children to reason and made it clear to them that we were interested in their reasoning rather than in their calculation skills.

We began the sessions on **randomness** by emphasising the everyday importance of randomising in activities such as shuffling cards and throwing dice in board games. One of our aims was for the children to discuss with each other the effects of randomising, and so they shuffled one pack of Happy Families and left another in its primordial state (i.e. grouped in families) and then we asked the children "*From which pack should you pick a card to be sure to get a female/child with the first pick? From where in the pack do you want to pick?"* The children thought about ways in which outcomes are randomised and wrote about their ideas on the uses of randomisation.

The dice-throwing activities involved fair and loaded dice, which created something of a sensation among these otherwise quite sophisticated 10 year olds. We asked the children about the predictability of the results of the next throw of a die (much more predictable in the case of the loaded dice) and our activities also involved discussions of the link between fairness and randomising. In later sessions we encouraged them to distinguish events that are possible though highly unlikely (*catching a fly with chopsticks*) from those that are entirely impossible (*catching* *a shadow*), our aim being to help them to understand that events that seem unlikely but are possible should be part of the calculation.



Session 1 Families Cards

pack A ✓ or pack B ☐ Why? Because in Pack A We Know the order.

pack A ✓ or pack B ☐ Why? Because miss field is at the bottom.

pack A ☐ or pack B ☐ Why? ...

Why do we shuffle cards at the start of many games?

We Shuffle cards because if they are in order it's easier to predict and it makes the game fair and it's random.

Figure 1. A child's answer sheet in an exercise involving a comparison between a Happy Families pack (A) with suits arranged in the same order for members of the family (mother, father, son, daughter) in each suit. Pack B was shuffled. The children first say which pack they would go to if asked to find a particular character and then give reasons for shuffling the pack before a game of Happy Families.

The main purpose of the sessions on **sample space** was to show children how to work out and reason about all the possible outcomes in particular situations. They started from thinking about how different combinations of features define different objects, a form of thinking that they were familiar with, and extended this form of thinking to defining different events. Figure 2 illustrates the use of tree diagrams, a tool that they learned about in our classes, in order to define different cars. We were pleased by how quickly pupils saw the point of these diagrams, how much they enjoyed creating them, and how easily they extended their use to the solution of probability problems.

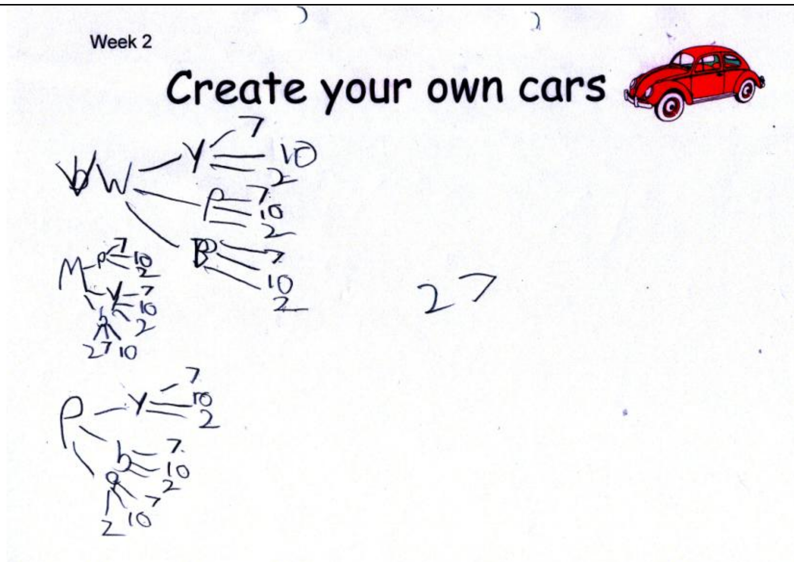


Figure 2. An example of a successful tree-diagram drawn by a pupil trying to find out how many different kinds of cars there could be with three possible makes of car (VW, Peugeot and Mercedes) in three possible colours (yellow, pink, blue) and with three possible numbers of seats (2, 7, 10). The correct answer is 27.

Over sessions, these combinatorial tasks became more sophisticated; for example, the children had to avoid duplications of each combination. Figure 3 gives examples of their ingenuity in working out all the matches that could be played between a number of teams.

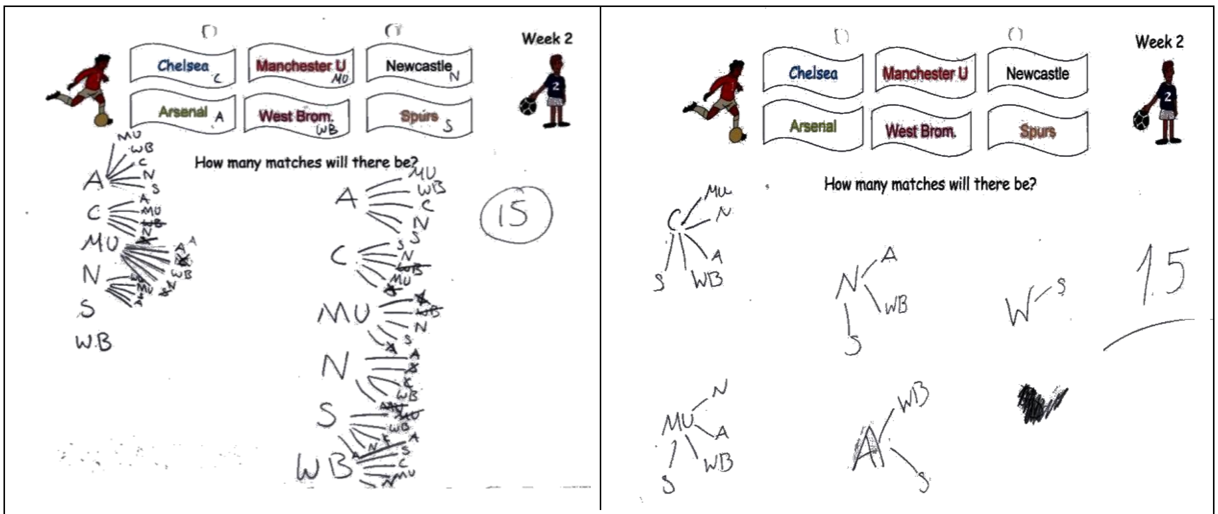


Figure 3. The pupils are asked to work out all the possible games of six football teams. They need to avoid is making the same pairing twice (Man United vs Chelsea is the same pairing as Chelsea vs Man United). The child on the right anticipates the repetitions and eliminates them during the production of the tree diagram. The child on the left first puts in all the combinations and then crosses out the repetitions in the diagram.

In the sessions on **quantification of probability,** the main theme was the use of ratios to analyse and to compare probabilities, as in the example in Figure 4. The figure shows how ratios can provide a striking way of showing children that calculations of probability are based on the relations between quantities and not on single quantities: there are different numbers of Dr Who fans in the two classes but the probabilities of meeting a Dr Who fan is the same in each class because the ratio between the two kinds of fan is the same.

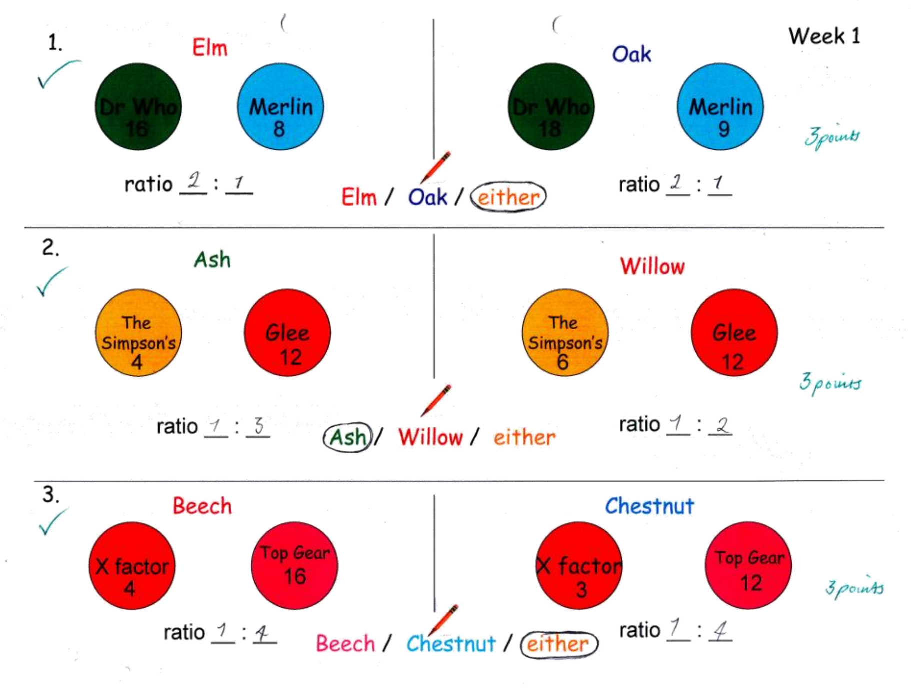


Figure 4. This is a simple and effective way of introducing ratios as a way of quantifying and comparing probabilities. Each question gives the relative numbers of fans in two different classes of two well known TV series, and the child is asked to work out in which class you would be more likely to meet a particular type of fan (e.g. a Dr Who fan in the first question). In the first and the third questions, the ratios between the two numbers of fans are the same in the two classes. In the second question, the probabilities are different (you are more likely to meet a Glee fan in Ash class than in Willow class)

The Effectiveness of the Probability Teaching Programme

Our prediction was that the children who took part in the probability teaching programme would respond at roughly the same level as the other two groups in their answers to the probability questions and to the problem solving questions in the opening pre-test, but would do better than the other children in their answers to the probability items in the immediate and in delayed post-tests. To test this prediction, we looked at the children's responses to questions on randomness, on sample space and on quantification separately.

The main statistical method that we adopted was *analysis of covariance*. We examined the differences between the different groups in the post-tests, and at the same time controlled for the effect of differences between the children in these groups in the pre-test: in technical terms the dependent variable was the children's post-test scores and the covariate was their initial performance in the pre-test.

Understanding of Randomness

As we predicted, the children in the Probability group made more progress in learning about randomness than those in the other two groups. In the pre-test (Time 1) the mean number of correct answers to the randomness questions was much the same for the three groups but the Probability group's mean scores were clearly better than those of the other two groups in the immediate post-test (Time 2). In a delayed post-test (Time 3, which was given about three months after the teaching on randomness had finished) this group still held the lead, as Figure 5 shows. So, these scores suggest quite strongly that the part of the probability teaching programme that dealt with randomness was effective: the children who went through the programme learned well about the consequences of randomness and the process of randomisation and remembered much of what they had learned over a long period of time. The difference between the groups' scores in the two post-tests was highly significant in an analysis of covariance that controlled for the differences in their initial scores in the pre-test ($F(2,75)=6.19$, $p=.003$).

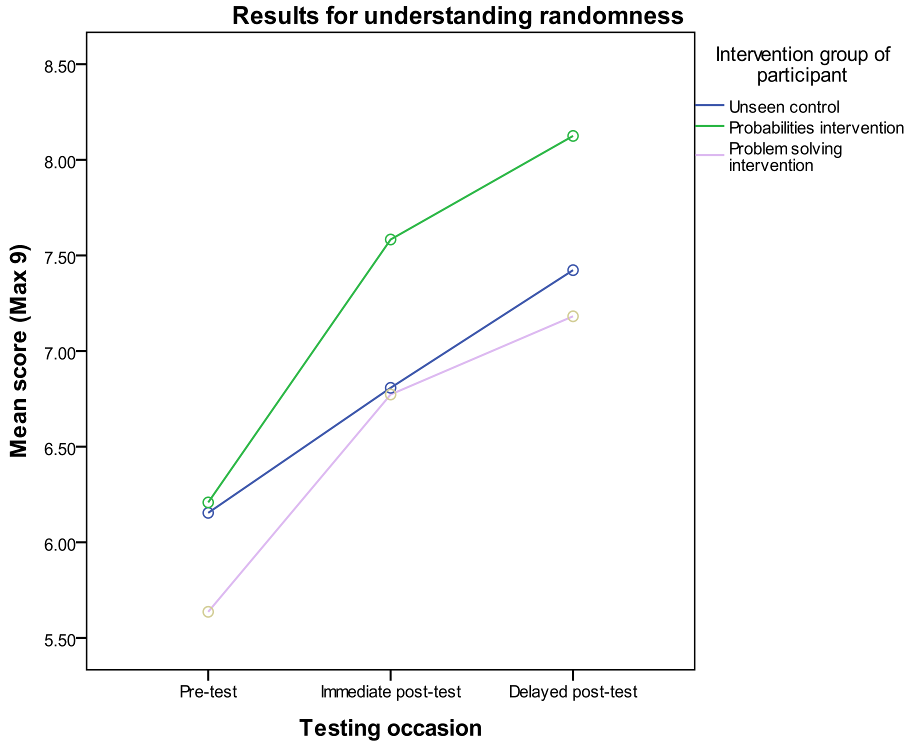


Figure 5. Mean scores in the assessment of randomness at the three testing occasions by group

We also asked the children to justify their answers to two of the questions on randomness in writing in the pre-test and in the two post-tests. These two problems were both about a person who was about to pick a marble at random from a bag which held marbles of two different colours (A & B) that had been randomly mixed: the numbers of marbles of each colour were specified (in one item there were more marbles of one colour than of the other and in the other item the number of marbles of the two colours was the same). The story was that this person had already picked and then replaced four marbles and these had all been of Colour A; each time that the chosen marble was replaced, the bag was shaken and the marbles were randomised. The question that the children were asked was about the colour of the person’s fifth pick: was colour A more likely than colour B, or colour B more likely than colour A, or were they equally likely? To answer the question correctly, the children had to realise that the colour of the marbles picked the first four times was irrelevant to the next pick: since the marbles were mixed up randomly when the bag was shaken before each pick, the only determinant of the probability of the next pick was the relative number of marbles of each colour in the bag.

We scored each justification by giving the child a point for mentioning any of the following aspects of each problem:

- uncertainty *"you don't know what will happen it is just luck"*
- randomness through mixing *"because the marbles are mixed up*"
- equality/inequality of the two colours *"it is because there is the same number of colours"*
- replacement *"because she put the marble back in the bag"*
- the run of four marbles of the same colour was just a chance happening *"it may have been just luck 4 purple marbles".*

The justification scores confirmed the growing understanding about the nature of randomness in the children in the Probability group, whose justifications in the post-tests far outstripped those of the children in the other two groups, as Figure 6 shows. An analysis of covariance confirmed that there was a clear and significant difference between the relevance of their justifications and that of the children in the other two groups (F(1,67)=6.43, *p*=.003.

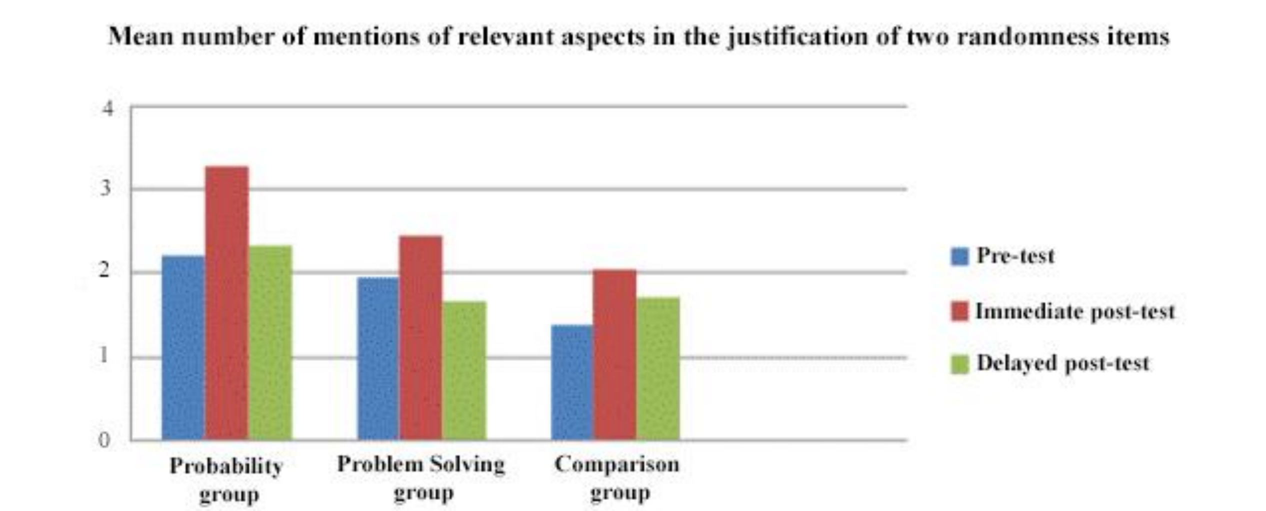


Figure 6. Mean scores in the justifications given for two randomness items by group

In the post-test items on sample space, the children were asked to generate all the possible combinations of properties that would define a particular event. In one of these items the children were asked to construct a tree diagram before they answered the question.

Working Out the Sample Space

Figure 7 gives the mean sample space scores of the different groups in the two post-tests. At the immediate post-test (Time 4), the Probability group already performed better than the other two groups and maintained this lead in the delayed post test, which was given 2 months (Time 5) after the teaching on sample space had been completed. This difference between the groups was significant in an analysis of covariance of the sample space scores in the post-tests (F(2,62)=9.92, *p*<.001).

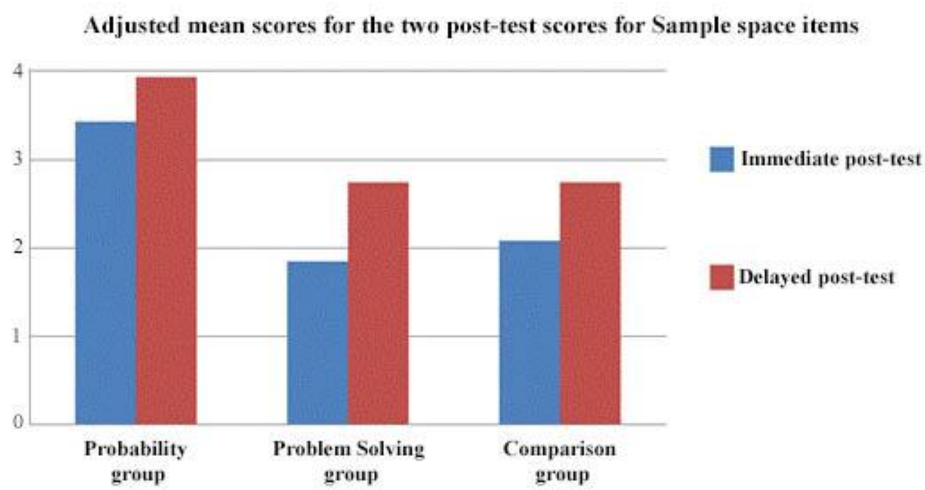


Figure 7. Adjusted mean scores (controlling for pre-test differences) on the sample space items

Quantifying Probability

We tested the children's ability to quantify the probability of two events at pre-test, immediate post-test and delayed post-test. The children were asked to indicate from which of two boxes they would rather draw a card, if the winner in the game was the one who pulled out a card with a circle. For example one box had 2 cards with a circle and 4 with a square, whereas the other box had 3 cards with a circle and 6 with a square. In this case, the probability is the same and the answer should be that it does not matter from which box they pull out a card, an option clearly explained to them at the time of testing.

The children in the different groups did not differ in their performance in the measures of understanding randomness and sample space in the pre-test at the start of the study, when they were still in Year 5. They did not take a pre-test in quantifying probability at that time. The quantification pre-test was given in the beginning of the Spring Term (Time 3) when they were in Year 6, just before they started to learn about how to measure the probability of an event. Figure 8 shows the three groups' mean number of correct answers to the quantification of probability items at times 3, 4 and 5.

Although the specific teaching had not started yet at Time 3, the Probability group performed slightly, but not significantly, better than the other two groups in this pre-test. By learning about randomness and sample space, the children in this group may have already started to develop an understanding of how probability should be quantified. Figure 8 also shows that the initial superiority of the Probability group continued in the two post-tests (Times 4 and 5) after the children in this group had been specifically taught about quantifying probability. One can see in the figure that the Probability group's performance did not decline between Test 4, given soon after they completed the teaching, and Test 5, given about 2 months after the teaching programme had ended. However, none of these differences between the children in the Probability group and those in the other two groups was significant in an analysis of covariance of the quantification scores in post-tests ($F(2,72)=1.92$, $p=.54$), and therefore it is not possible to claim that they are due to the effectiveness of our teaching. For this reason, we used the extensive protocols of the children's solutions in this part of the intervention to redesign our intervention on quantifying probabilities in order to make it more effective in the second phase of the project.

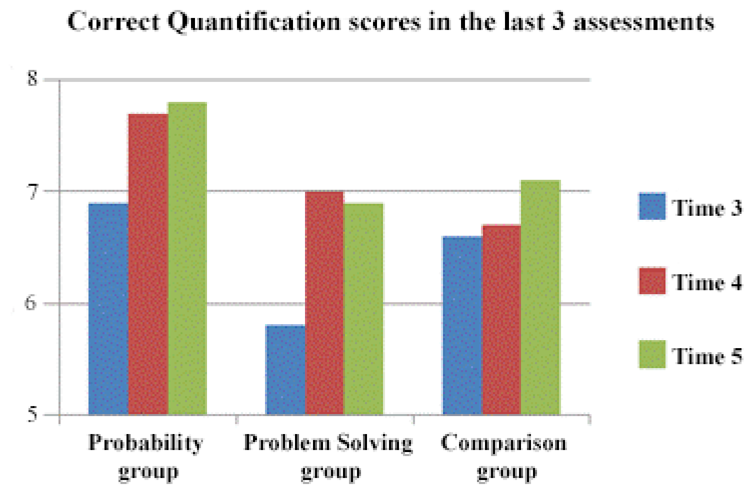


Figure 8. Mean scores in the quantification of probabilities items by group at different times

Summary

The three parts of the probability intervention programme in the first phase met with mixed fortunes. We concluded that we had developed a successful programme for teaching children about randomness and the sample space since the children in the Probability group learned significantly more about these important probability concepts during the first two terms of the first phase than the other two groups. We felt confident therefore that teaching can play a decisive role in helping children to understand these two probability concepts. Our results on items about quantifying and comparing probabilities were not so strong, which we took as demonstrating that this part of the intervention needed to be adjusted and improved.

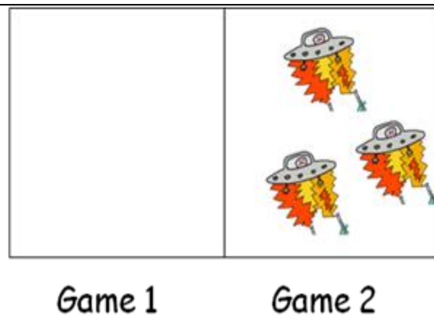
We concluded that some of the activities in the unit on teaching children about quantifying probability had proved too difficult for children in the 9 to 11 year old age group, and also that some of the items that we had used in the assessment of the children's understanding of quantities were also too hard, and, in accordance with the design experiment approach, we adjusted the intervention and the assessment accordingly for the next phase of the project.

TEACHING PROBLEM SOLVING

The Problem Solving Teaching Programme in the First Phase

The problem solving programme had two units: the first was designed to help children to think about the inverse relations between operations and the second focused on relations between quantities. As in the probability programme, the children worked in pairs: each child solved the problems individually, then discussed the solution with one peer, and finally the pair's solution was presented to the group. During the second unit, each pair produced an acetate and projected their solutions using an overhead projector to support their presentations.

The unit on **inverse relations** between operations contained problems that involved either natural or directed (positive and negative) numbers. The latter problems were about games in which the players could win and lose points. The children calculated the scores of successive games, which could be positive or negative. The language used was "points won" or "lost", as previous research (Nunes, 1993) has shown that children deal better with directed numbers when they do not use the formalisation of plus and minus signs. During this unit, the children were encouraged to use red and yellow cards to represent points won or lost, if they were unable to calculate scores. An example of this type of problem is illustrated by the Gremlin's game (see Figure 9). The children were told that a point is scored for each Gremlin that is hit and a point is lost for each spaceship that is hit. The game outcomes are presented on a screen; the Gremlins and spaceships that were hit disappear from the screen. The children filled in the scores in their answer sheets. After they were comfortable with calculating scores, they were presented with a sequence in which the score for the first game was missing but the final score was known. In order to calculate the score for the first game, they needed to consider the inverse of each game's score. Figure 9 shows an example in which the two slides in the series are presented simultaneously; no Gremlins appear, only the spaceships that were hit. A snapshot of the children's answer sheet is presented below the slides.



What happened in game 1? ________	Score for Game 2 ________	Final score __won 5___

Figure 9. An example from the Gremlins game

Even children who had shown no difficulty with the inverse relation between addition and subtraction as inverse operations in the context of natural numbers had to pause when facing the type of problem exemplified in Figure 9. As one of them asked when trying to invert the operations to return to Game 1: "*What is minus minus 5*?"

Many children had difficulty when a negative score in the second game led to a final positive result, but some children could explain their reasoning in their own words, which helped those who were struggling. For example, one child said: "*You have to put back the points that he lost to find out what he scored in the first game, so if he has 3 and lost 5, and you put the 5 back, then you know he had 8 before.*"

The inverse relation between operations was the focus of the seven initial lessons, but this knowledge continued to be relevant because inverse relations are often part of the reasoning in solving problems. Figure 10 shows the explanation provided by one pair of children during a discussion of the problem: “Helen had some sweets. She gave half to her friend James and 3 to her cousin Alex. Her grandfather gave her 20 more the next day. At the end she had 23 sweets. How many sweets did she have at the beginning?"

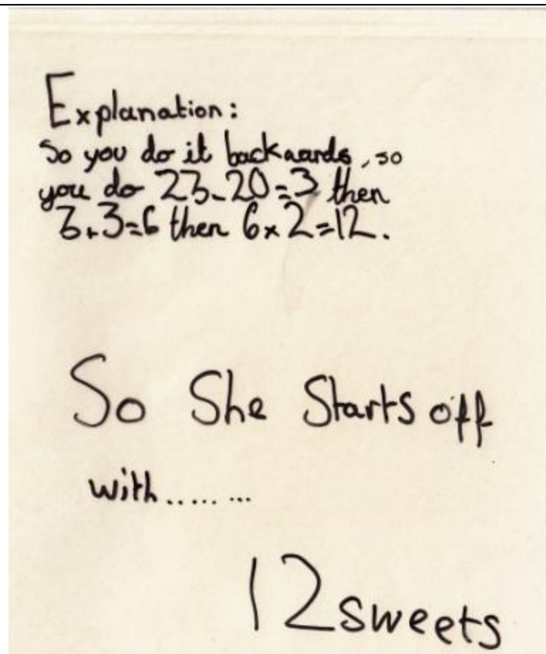


Figure 10. One pair's explanation of how they found the solution

The second unit focused on **exploring relations between quantities** before solving a problem. Sometimes the quantities were unknown, and only the relations were mentioned in the problem (see the problem about Billy and Jason presented in the introduction). Sometimes there were known quantities but it was necessary to consider relations between quantities before solving the problem (see the problem about Kate, Donna and Jamie sharing stickers in the introduction). Some problems involved additive reasoning and others involved multiplicative reasoning.

We introduced bar diagrams to represent additive relations between quantities. Figure 11 shows the first problem in Unit 2 and illustrates the introduction of the bar diagram used by the researchers after the children had solved the problem. The children answered either 5 or 10; the majority answered 5.

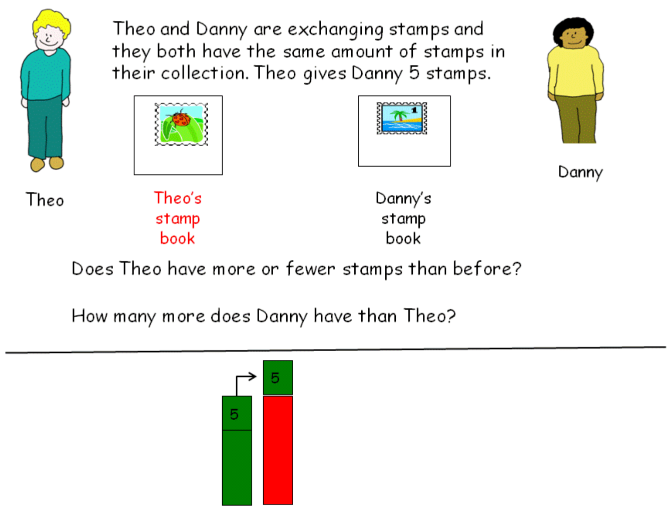


Figure 11. The first problem in Unit 2 and the bar diagram used to discuss the solution

The researcher then presented the bar diagram on the screen and asked the children to think about the problem again. When the segment marked 5 was moved from Theo's to Danny's stamp collection, the children immediately recognised that the difference was then 10 stamps. The researcher suggested to the children that they should use bar diagrams to help them solve the subsequent problems. This allowed the children to create their own diagrams and discuss their solutions with the group by presenting their own diagrams or explanations using the overhead projector.

Figure 12 shows the diagrams produced by two different pairs for the problem: "On sports day, Year 4 won 3 prizes more than Year 5 and Year 6 won 5 prizes more than Year 4. The school gave out 68 prizes. How many prizes did the champion class win?"

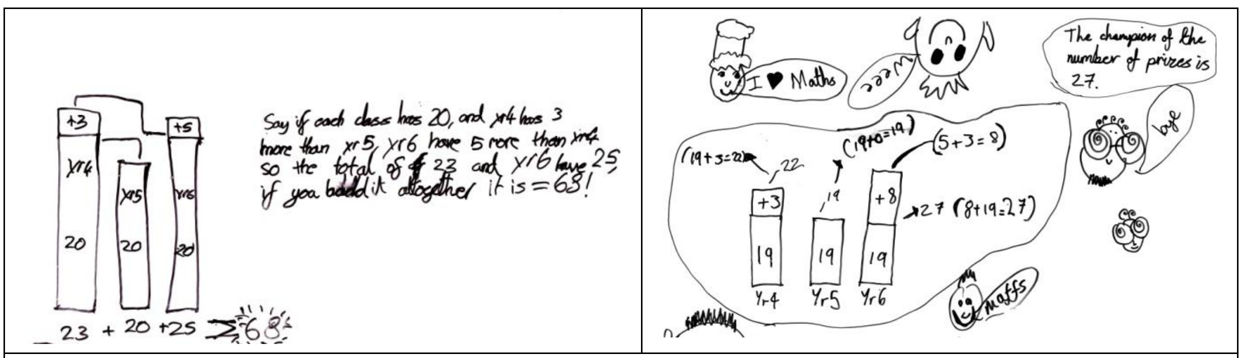


Figure 12. Two pairs of children produced different solutions for the same problem

The diagram on the right shows the composition of relations for Year 6: 5 more than year 4, which had won 3 more than Year 5, means that Year 6 won 8 more prizes than Year 5. The diagram on the left does not show this composition of relations; it indicates that Year 6 won 5 prizes more and a line suggests that the children read the problem correctly because it connects the relation "+5" to the bar representing Year 4's prizes. During the discussion, the children who had produced the diagram on the left realised that they needed to compare the prizes that Year 4 and Year 6 won to those won by Year 5 in order to solve the problem.

Adding relations when the quantities are unknown was a persistent difficulty for many children throughout the programme. On the last day of the intervention, this problem was presented: "James, Kelly and Donna were playing a computer game. At the beginning they had the same number of points. James played against Kelly and lost 20 points which went to Kelly. Then, Kelly played against Donna and Kelly lost 30 points which went to Donna. What is the difference in points between James and Kelly now? What is the difference in points between Donna and Kelly? What is the difference in points between Donna and James?" Most of the pairs solved this problem by attributing an initial number of points to the players and calculating how many points they would have after playing the games. Thus they operated on quantities (number of points) rather than relations and then compared the quantities. Figure 13 illustrates this type of solution. The solution was used in a fruitful discussion: the researcher asked the children whether a different solution would be found if a different starting point were used (e.g. 60 points) and whether it was possible to solve the problem without attributing a number of points to the players at the start. These questions promoted quite a lot of discussion.

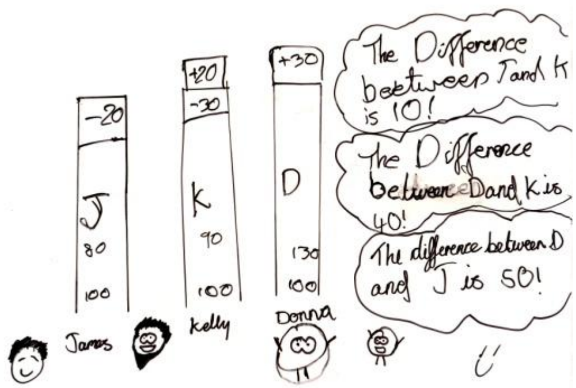


Figure 13. The same value (100 points) is attributed to the players

Bar diagrams were used to support the solution and discussion of additive reasoning problems, but we chose line diagrams for the representation of problems involving proportions because they offer the possibility of establishing correspondences between values in the different measures. Figure 14 presents the problem used to introduce the line diagram, which was adapted from Streefland (1984). In the discussion of this problem, we also introduced the use of doubling and halving as strategies to find answers; for example, when one doubles the number of steps that Mrs Elastic takes, one should also double the number of steps that Dazz takes. We also discussed with the children that they could find easily how many steps Dazz takes to keep up with Mrs Elastic when she takes 3 steps by halving the numbers. Finally, we asked them how they could use this information to find out how many steps Dazz took to keep up with Mrs Elastic when she took 15 steps. This discussion draws on the children's well documented intuitions about proportions: they understand that if they double the value in one quantity, they must also double the value in the other quantity (e.g. Hart, 1981; Schliemann & Nunes, 1990; Singer, Kohn, & Resnick, 1997). The use of parallel operations on each of the measures is known as "within measure" or "scalar" solution. The children had the opportunity to solve some further problems in this way.

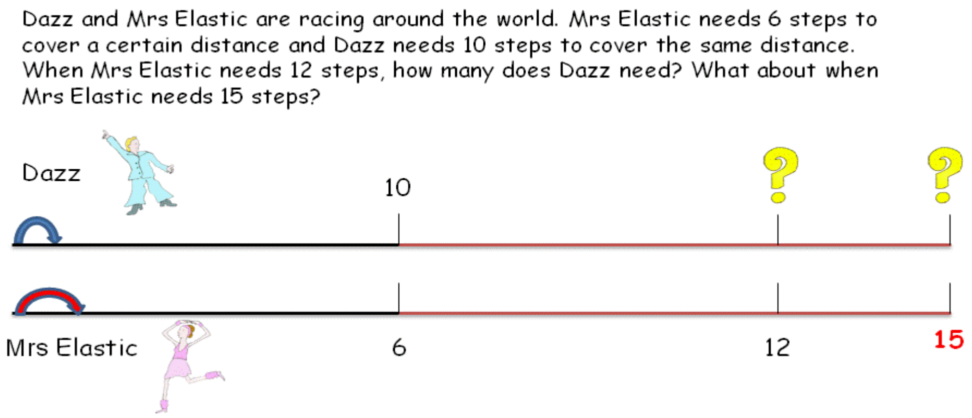


Figure 14. The problem used to introduce the ratio lines

Figure 15 shows one pair's solution to the problem: “Jim has to print the school newspaper but he can only do it in break time. It takes him 15 minutes to print 12 newspapers. How many newspapers can he print during the 35 minutes break**?”** The children in this pair first doubled the initial values to find how many newspapers can be printed in 30 minutes; then they realised that Jim had 5 minutes left, which is 15 (minutes) divided by 3, and calculated how many papers are printed in 5 minutes; finally they added the number of papers that can be printed in 5 minutes to the number that can be printed in 30 minutes.

Our aim was to further the children's understanding of proportions by helping them to realise the usefulness of functional solutions, which involve operations between measures.

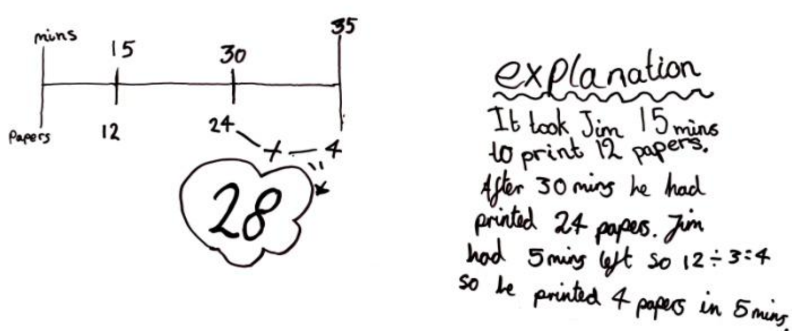


Figure 15. The use of scalar relations in solving a proportions problem

Because the functional relation is constant, it provides a useful insight into proportional relations. Figure 16 shows an example of a problem in which we facilitated a discussion about the functional relation between the variables. The problem was: “Granny has to put fertilizer on her plants in her garden. The instruction says that she has to mix 2 measures of fertilizer with 8 litres of water (the measure comes in the box). She has a tank with 24 litres of water in it. How many measures of fertilizer does she have to put into her tank?”

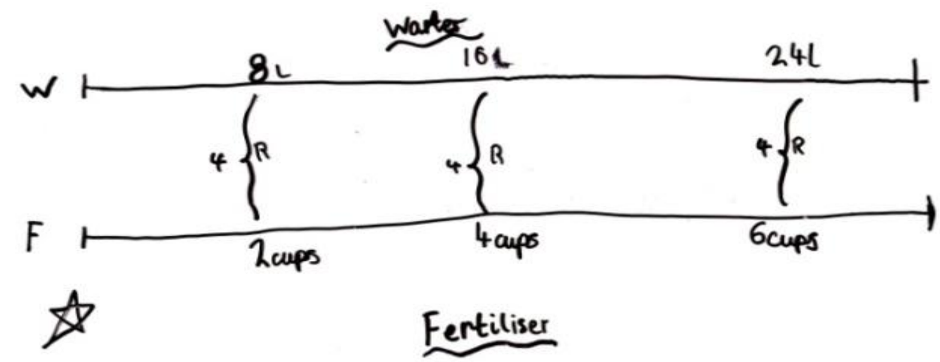


Figure 16. A solution with the marking of the constant relation between the quantities

Although the children recorded the functional relation and also spoke about the fixed ratio of one quantity to the other, they continued to prefer scalar (or within quantities) solutions. The concept of functional relation remained a difficult one.

We also noted that the use of the line diagram did not prevent additive solutions to problems involving proportions. The problem in Figure 17, left, is often used in research about proportional reasoning. Children's preferred solution to this problem is additive. We included it in our programme and found that most solutions were additive, even if the children had used the line diagram. Figure 17, top right, shows a diagram demonstrating an additive solution. The relations indicated in the diagram suggest that the children used the idea of a fixed relation between the quantities but conceived of it as additive rather than multiplicative.

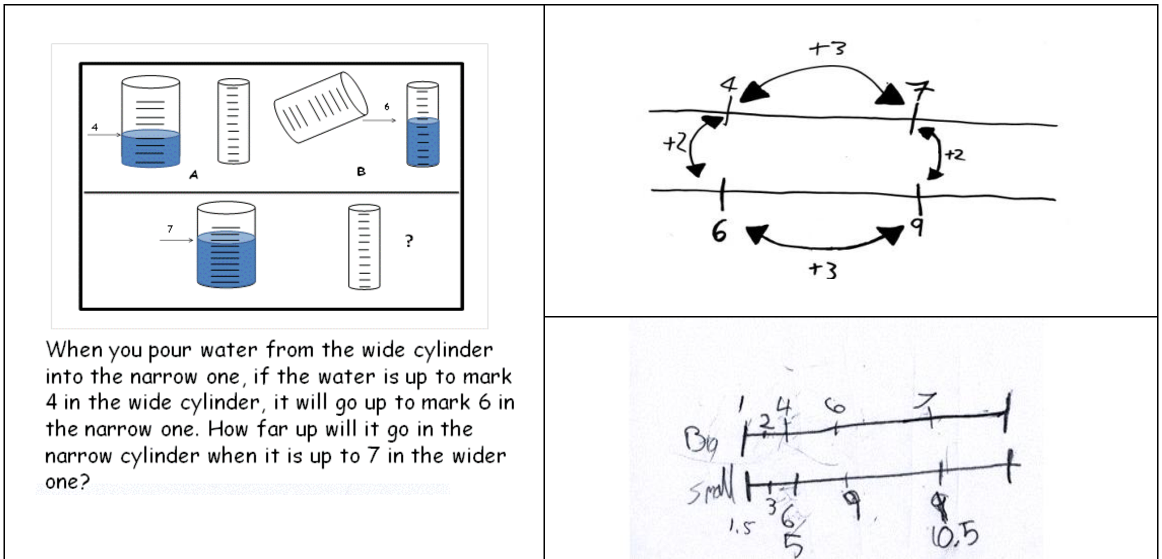


Figure 17. A ratio problem and two diagrams used to solve the problem

The bottom, right diagram shows a solution by a boy, who originally used an additive solution. Prompted by the researcher, he proceeded to halve the initial data to check his solution. By doing this, he realised that if the water was up to 2 in the wide cylinder, it should be up to 3 in the narrow one, and if the water was up to 1 in the wide cylinder, it should go up to 1.5 in the narrow one. He then established by trebling 2 in the wide and 3 in the narrow cylinder that, if the water was up to 6 in the wide cylinder, it should be up to 9 in the narrow one. This led him to realise that his original answer was wrong, and he corrected it to 10.5. When his turn came to explain his solution to the group, he found it difficult to articulate why he thought that the multiplicative solution was a better one.

The researcher running the session did not explore the problem further. However, it would have been possible to explore the additive reasoning diagram in order to create lively discussions with the children regarding the appropriateness of an additive solution. By applying additive reasoning, the children should subtract 4 from the marks on the wider and on the narrower cylinder; this would show them that, by their calculation, when the water was up to mark 2 in the narrower cylinder, the same amount of water would be at zero in the wide cylinder. We wonder whether the children would see the inconsistency of the additive solution and reject it.

Our reflections about the use of the diagrams during this phase suggested to us that the use of diagrams is not akin to the use of a procedure that can be followed without considering the meaning of the problem: in fact, the child's interpretation of the problem is reflected in the way the diagram is drawn. Children make mistakes when they use the diagrams, but the discussion of the diagrams produced by the children can clarify the reason for obtaining different solutions and can be used to further the children's understanding of relations between quantities. Such discussions can be seen as dialogues in the zone of proximal development (Murata, 2008) and are probably essential for the successful implementation of diagrams as problem solving tools. They require the researcher or teacher to be very familiar with the problems as well as with the children's probable mistakes.

Although the problems had been taken from research with or books for primary school children, our observations during the lessons indicated that some problems were too difficult for the children because none of the pairs of children arrived at a correct solution. These difficult problems took large chunks of time and seemed unproductive because the children were less active in the discussion when the researcher had to take the lead in drawing a diagram that provided the basis for a correct solution.

The Effectiveness of the Problem Solving Teaching Programme

During this phase of the project, we designed and evaluated our assessments of the children's achievement in problem solving. Similarly to the procedure for assessing probability concepts, we used different assessments at different time points, both to keep up with the children's mathematical progress during the year and to obtain better measures, as we evaluated what had been used at different time points. Items that were found to be too easy or too difficult were not included in later assessments.

Understanding the Inverse Relation Between Operations

Our prediction was that the children who took part in the problem solving teaching programme would respond at roughly the same level as the other two groups in the opening pre-test, but would do better than the other children in the immediate and in the delayed post-tests of problem solving. To test this prediction, we looked at the children's responses to the questions that involved reasoning about the inverse relations between operations that were presented in the context of games or in the context of equations (e.g. c-25=8; what is c?; a-15=-6; what is a?).

Figure 18 presents an overview of the results at pre-test, the immediate post-test (given after the five lessons in the first term of the project, which was a Summer Term) and the delayed post-test (given at the beginning of the Autumn, when the children were starting the new school year). These tests were exactly the same. The groups differed to some extent at pre-test; the Comparison group performed better than the other two groups. At the immediate post-test, the Problem Solving group had caught up with the Comparison group and at the delayed post-test it overtook the Comparison group. In order to assess whether the Problem Solving group had made more progress than the other two groups, we used an analysis of variance with repeated measures. Greater progress by one group in this analysis is indicated by a significant interaction between group membership and testing occasion, which shows that the differences in slope are significant. The interaction between group and testing occasion was significant ($F(2,69)=4.23$; $p<.05$). It is noteworthy that the Comparison and the Probability groups showed a decrease in scores, an effect known as the "Summer holiday effect", but the score for the Problem Solving group did not slide back and showed a small increase.

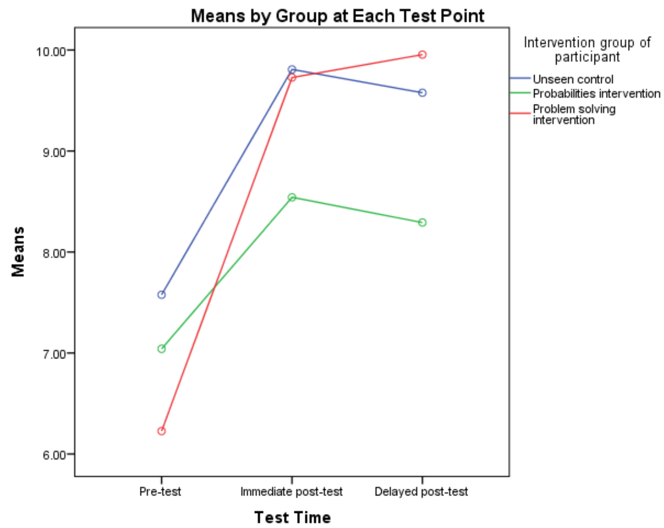


Figure 18. Means in the problems about inverse relations by group for each test time

A shorter version of this test was given again to the children at the end of the Autumn term, when the Probability group and the Problem Solving group had participated in five more lessons of their respective programmes. The means presented in Figure 19, left, are statistically adjusted for pre-test differences; the overall difference between the groups is significant ($F(2,73)=5.31$; $p=.007$). The Problem Solving group scored significantly higher than the Comparison group ($p=.003$) and the Probability group's score was also significantly higher than the Comparison group's ($p=.03$). Thus the Problem Solving group continued to perform significantly better than the Comparison group, even though some of their lessons were no longer focusing on the inverse relation between operations, and the Probability group overtook the Comparison group, although the children in the Probability group had no lessons about the inverse relation between operations. This result indicates that the Probability group was learning about concepts specific to probability problems, such as randomness and sample space, as well as learning about problem solving in a more general way.

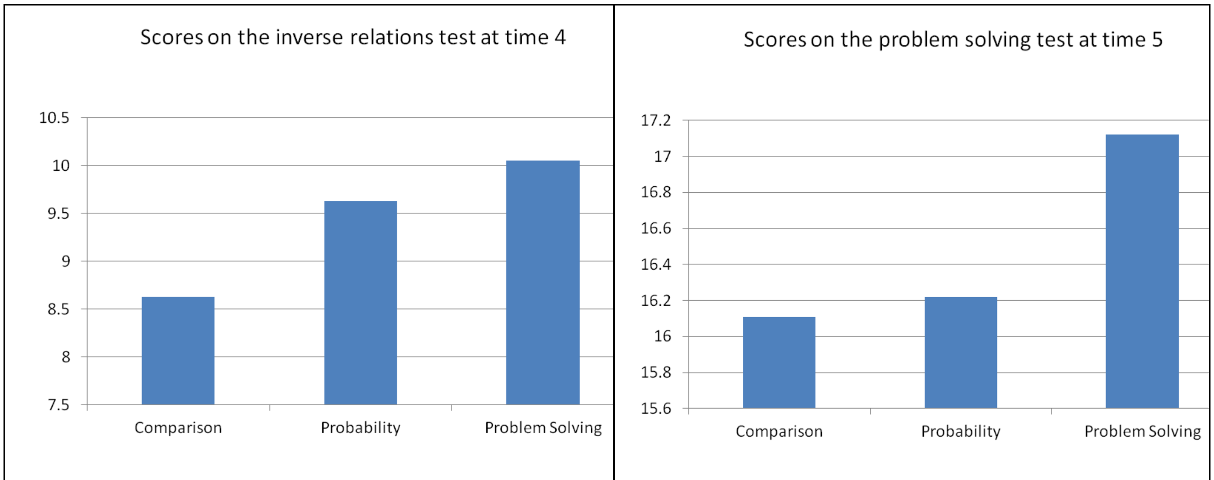


Figure 19. Adjusted means (controlling for pre-test scores) for Time 4 on the inverse relations test (left) and Time 5 on the problem solving test (right)

Understanding Relations Between Quantities

For the final comparison between the groups at Time 5, we designed a new test that evaluated problem solving skills in a more general way. It contained questions on games with positive and negative numbers, which had been used in the pre-test, no questions on equations, and new questions about additive and multiplicative problems that required the children to reason about relations between quantities. The latter problems had not been used in the pre-test. The adjusted means, controlling for pre-test differences, are presented in Figure 19, right. Although the trends in the means are similar to what was observed at Time 4 when we analysed the children's understanding of inverse relations between operations, the differences at Time 5 were not statistically significant, so there is not solid evidence that using the diagrams had helped the children to reason better about relations between quantities.

Summary

The Problem Solving group made more progress throughout the teaching programme than the Comparison group in assessments of their understanding of the inverse relations between operations. The difference between the groups persisted even after the emphasis in the teaching programme was no longer about inverse relations between operations. However, we did not find a significant effect of the use of diagrams on the children's reasoning about relations between quantities. We think that some of the problems we used were too difficult, leading to unproductive discussions, and removed these from the programme to be used by teachers in the second phase of the project.

THE SECOND PHASE: PROBABILITY AND PROBLEM SOLVING IN THE CLASSROOM

In the second phase, new and revised versions of the same two programmes were used by teachers in their own classrooms. Teachers were invited to participate and to attend a professional development day, which started with a discussion about the significance of mathematical reasoning for children's achievement in mathematics. They were given the opportunity to volunteer to use in their classrooms either the probability or the problem solving programme. A total of 43 teachers from 25 schools participated in the project. The teachers taught Years 5 or 6. Because the difference in year group was confounded with school differences (the Year 5 and the Year 6 children were in different schools), we could not study the effect of school year on the effectiveness of the programmes in this study. In nine schools, different teachers chose to participate in the alternative programmes. Although we had some concern about the teachers sharing information about their programmes, which could dilute the effect of the programmes, we agreed to their choices and asked the teachers not to share their experiences during the project.

The children who participated in this second phase formed two groups and each group acted as the Comparison group for the other.

1. *The Probability group* worked on probability problems
2. *The Problem Solving group* worked on non-routine mathematical reasoning problems.

The teachers were asked to follow the programme they had chosen as closely as possible, making allowances for differences in pace according to their children's response to the programme. They were asked to use the programme in one session a week for 15 weeks. However, some teachers decided to use the programme lessons more often than once a week and others less often.

The researchers administered a pre-test to each class before the teacher started the programme and a post-test when the teachers indicated that they had reached the end of the programme or at the end of the school year, even if the programme had not been completed. The interval between pre- and post-test varied because of the differences in the pace of the programme implementation among teachers. The interval between pre- and post-test varied from 3 to 8 months. Teachers from three schools were unable to schedule a post-test. A total of 783 children completed both tests; 430 participated in the Probability programme and 353 in the Problem Solving programme.

The different sections in the probability and in the problem solving assessments were significantly correlated over time, thus showing good test-retest reliability. A principal component analysis based on the scores for groups of similar items within each assessment showed that a single factor explained more than 50% of the variance in each assessment. Thus the assessments showed good construct validity.

Probability Results in the Second Phase

The probability intervention programme was designed to test the hypothesis that it is possible to teach the subject of probability successfully in primary school because it involves issues, such as fairness and predictability, that are inherently interesting – even fascinating – to primary school children. The next and crucial step was to find out if the programme would also be successful when delivered by teachers in their own classrooms. This was the purpose of the second phase of the project.

In most respects this second phase was the same or almost the same as the first phase. It involved the same teaching programmes for probability and for problem solving (both adjusted in the light of our experiences in the first phase) and the same (but also adjusted) pre- and post-test assessments. Again we used analyses of covariance as our main statistical tool for measuring the effects of the two programmes. There were, however, three radical differences between the two phases, apart from the fact that the programmes were delivered by class teachers in the second phase and not by us. These were:

1. There were only two groups in the second phase, each one acting as a control group for the other group. One group of children participated in the probability programme, and the other group in the problem solving programme.
2. The assessments were a pre-test and an immediate post-test only; these assessments were identical. The teachers did not have copies of these tests and did not know that the post-test would be identical to the pre-test. We were unable to organise a delayed post-test for practical reasons.
3. Although the teachers taught the children in the Probability group about all three of the basic aspects of probability, for practical reasons we were only able to assess the children's ability to work out the sample space and to quantify probability.

We found that the scores of the children in the Probability group improved from pre-test to post-test more than the scores of the children in the Problem Solving group in the items that tested their ability to analyse Sample Space, as Figure 20, left, shows. This difference was significant in an analysis of covariance of the post-test scores, that controlled for the children's scores in the pre-test ($F(1,892)=40.59$, $p<.001$).

Figure 20, right, shows that the scores for the items that tested the children's ability to quantify and compare probabilities improved in both groups from pre- to post-test, but this improvement was greater in the Probability group than in the Problem Solving group. The difference between groups was also significant in an analysis of covariance ($F(1,894) =7.49$, $p=.006$) . This positive result suggests that the adjustments that we made to our original intervention on quantifying probability were effective, and therefore that children of this age can learn successfully about quantifying probability, as well as about the other basic aspects of probability.

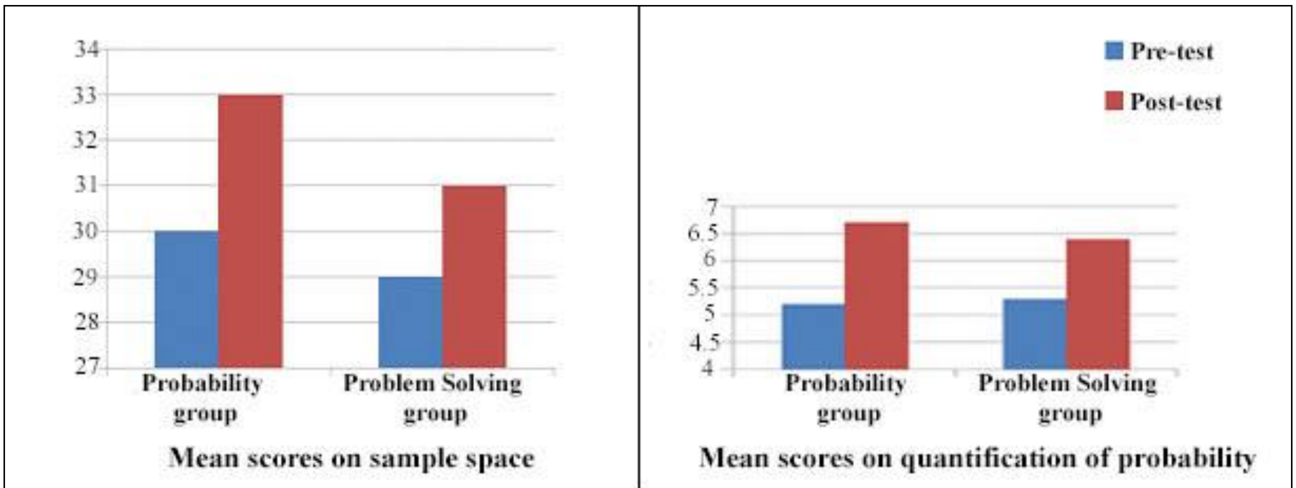
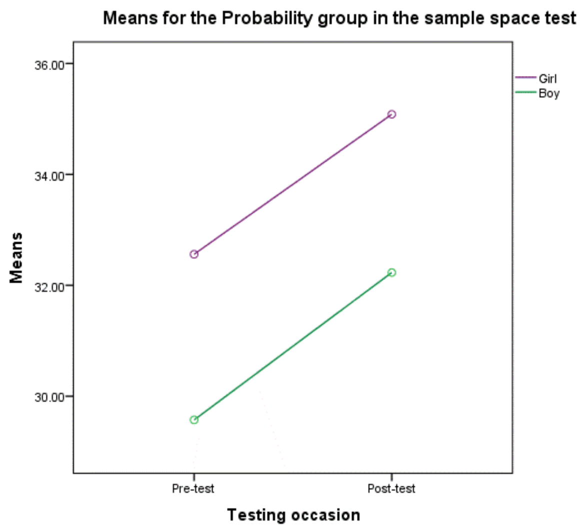


Figure 20. Mean scores on the measures of sample space and quantification of probability by group and by testing occasion

Previous research has suggested that boys and girls differ in some problem solving tasks. The large number of participants in this study allowed us to investigate gender differences in the probability measures as well as the effectiveness of the programme for boys and girls separately. Comparisons at pre-test showed an unexpected pattern: the girls performed significantly better than the boys in the sample space test whereas the boys performed significantly better than the girls in the quantification of probability test. The statistical analyses showed that both girls and boys in the Probability group did significantly better than the girls and boys, respectively, in the Problem Solving group in the post-test, controlling for the pre-test differences. The differences between boys and girls remained roughly the same after they participated in the teaching programme about probability. Figure 21 shows the pre- and post-test results for the Probability group by gender for each testing occasion.



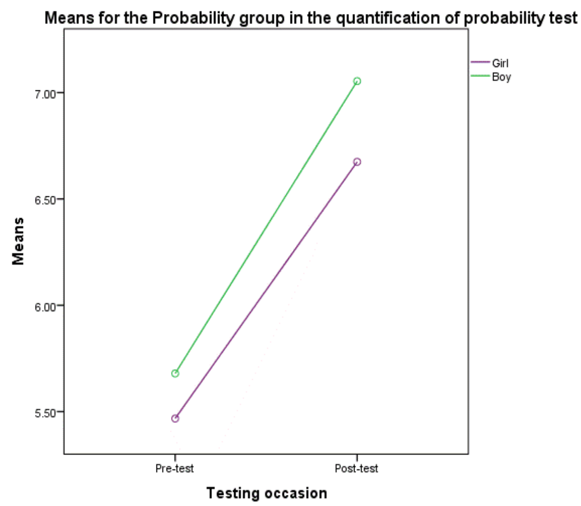


Figure 21. Means for the Probability group in the sample space and in the quantification of probability tests by gender and testing occasion

Thus the second phase provided further, vital evidence that it is quite possible to teach primary school children successfully about the concept of probability.

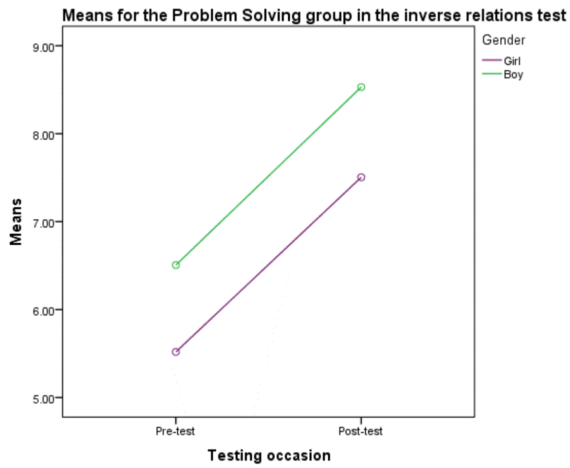
Problem Solving Results in the Second Phase

We analysed whether the problem solving intervention was effective when used by teachers in their own classrooms with pre- and post-tests similar to those administered in the first phase of the project. There were two scores: one assessed the children's understanding of inverse relations in the context of games or equations and the second was a general problem solving measure, which also included problems that required analysing the relations between quantities before implementing calculations (e.g. the problem about Kate, Donna and Jamie sharing stickers). The results of these analyses are summarised in Figure 22. In both comparisons, the two groups made significant progress from the pre- to the post-test (for the inverse relations measure, $F(1,808) = 42.77$; $p<.001$); for the general measure of problem solving, $F(1,808)=58.94$; $p<.001$). A significant interaction between testing occasion and group indicated that the Problem Solving group made significantly more progress than the Probability group in the measure of understanding inverse relations ($F(1,808)=14.29$; $p<.001$) as well as in the general measure of problem solving ($F(1,808)=6.37$; $p=.012$).

Figure 22. Means at pre- and post-test in the inverse relations and problem solving tests by group

Previous research has found differences between boys and girls in problem solving tests. As the number of participants in this phase of the project is large, it was possible to analyse gender differences. We found that boys performed better than girls in the pre-tests, and so we analysed the results of the intervention by group and by gender. We wanted to know whether the programme was more effective for boys or girls or whether it was equally effective for both groups. It was possible for the programme to be equally effective for both but the gap between boys and girls in problem solving not to change.

The statistical analyses showed that both boys and girls in the Problem Solving group improved significantly from pre- to post-test and that they improved more than the boys and girls in the Probability group (the analyses show a significant effect of group, reported earlier on; the interaction between group and gender is not significant). Figure 23 shows the comparison between boys and girls in the Problem Solving group at pre- and post-test. As the parallel lines in each graph indicate, the effect of the programme did not differ for boys and girls: they show similar levels of improvement from pre- to post-test. In the inverse relations measure, there is a significant effect of gender ($F(1,808)=28.39$; $p<.001$); the interaction between gender and group is not significant. In the problem solving measure, there is also a significant effect of gender ($F(1, 808)=10.32$; $p=.001$); the interaction between gender and group is not significant). This means that the programme did not close the gap between boys and girls but that the girls profited as much as the boys from it.



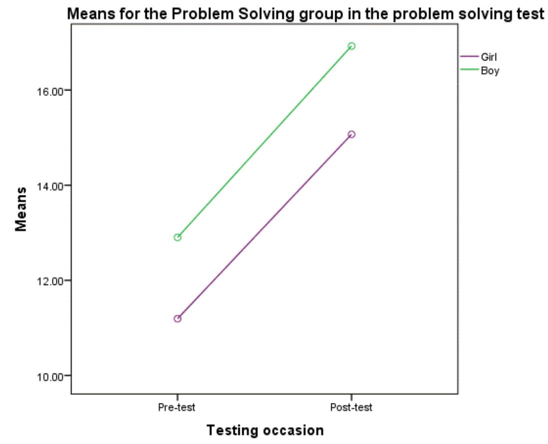


Figure 23. Means in the two measures used to assess the effectiveness of the problem solving intervention for boys and girls

In summary, the teachers were able to implement the problem solving programme successfully and the children who participated in the programme performed significantly better than those in the Probability group in the problems that do not involve probability concepts. The Probability group progressed significantly in the problem solving assessment from the pre- to the post-test, but not as much as the group that was specifically taught about problem solving. We conclude that this project provides crucial evidence that it is possible for teachers to teach problem solving in meaningful ways in primary school.

DISCUSSION AND CONCLUSIONS ABOUT THE WHOLE PROJECT

The underlying theme of this Nuffield-funded project was the need to find effective ways for teachers to encourage and improve their pupils' mathematical reasoning. Our previous research had shown that children's ability to reason about relations between quantities plays an important part in their learning of mathematics even in their first years at school (Nunes et al., 2007). This new project was about teaching children how to deal with mathematical reasoning problems at the end of primary school.

Every mathematical reasoning problem that children are asked to solve involves thinking about relations between quantities in the problem. This relational reasoning is exceptionally important in probability problems, which require pupils to recognise randomness and to work out systematically what is called the sample space. Probability problems are not part of the new primary school mathematics curriculum in England, but this project began with the idea that this might be an unnecessary and an unfortunate omission. So, two separate programmes were developed for teaching mathematical reasoning. One was designed specifically to teach children about probability, and the other to teach children about mathematical reasoning in problems that do not involve probability. The purpose was not only to establish a successful form of instruction of mathematical reasoning in both these contexts but also to find out whether probability is a subject that can be taught to primary school children in a comprehensive and a comprehensible way.

The overarching result of the project was that the two teaching programmes, one for solving probability problems, the other for problems which did not involve probability, were successful in an interesting and rather specific way. In the first phase, the design phase, the children who were taught how to reason about problem solving became much better at solving problems that do not involve probability than those taught about probability and those who went through neither programme. In much the same way, the children who joined the probability programme produced better scores in the post-test assessments of their answers to questions about randomness and sample space than the other two groups of pupils did. The probability programme did not produce a significant difference in the groups' ability to quantify and compare the size of different probabilities, but this failure was attributed to details of the teaching programme, which was adjusted before the second phase

Of course, much of the success of the two programmes in the first phase could have been due to our own zeal, since we were self-confessed enthusiasts for teaching children about mathematical reasoning in general and for introducing children to the reasoning that is needed to solve probability problems at an early stage in their school career. However, when, in the second phase, the two programmes were handed over to teachers to carry out in their own way and without direct supervision from the researchers, the results produced by each of the programmes were either very similar to those in the first phase or actually more positive than the first phase results.

The children in the Problem Solving group did better than the children taught about probability in the problem solving questions in the post-test assessment of reasoning, and the children taught about probability were able to reason significantly better than those in the Problem Solving group about the probability problems given to them in the post-test assessment.

The educational implications of these results are clear and simple.

1. One important conclusion drawn from the project is about teaching mathematical reasoning in general. The method of giving pairs of children problems to solve together and then encouraging them to discuss their joint solutions with the rest of the group worked extremely well, and apparently worked as well with one kind of reasoning as with the other.
2. For the first time it has been established that 10 and 11 year old children can learn a great deal in the classroom about all three of the fundamental aspects of probability. This means that our programme can serve as a model for a coherent introduction to the understanding of probability in primary school classrooms.

3. It has also been established for the first time that primary school children’s skills in problem solving improve significantly by learning explicitly about the inverse relation between operations and by learning to use diagrams as tools for solving and discussing mathematical reasoning problems.
4. This study has also showed that, although the children's performance in the problem solving and in the probability measures are correlated, the outcomes of teaching mathematical reasoning in and out of the realm of probability are independent: children can make progress in one but not the other type of reasoning. This strongly suggests that leaving primary school children without instruction on probability leaves them unprepared to learn about probability later, even if their general problem solving skills have developed significantly.
5. The teachers’ and the researchers’ informal experiences in the study was that the children who took part showed a great interest in and liking for both forms of mathematical reasoning and enjoyed the closing group discussions of the problems that they had been given and had tried to solve.

REFERENCES

Brown, A. L. (1992). The Design Experiment: theoretical and methodological challenges in creating complex intervention in classroom settings. *Journal of Learning Sciences, 2,* 141-178.

Bryant, P. & Nunes, T. (2012) Children's understanding of probability. A literature review (full report). http://www.nuffieldfoundation.org/sites/default/files/files/Nuffield_CuP_FULL_REPORTv_FINAL.pdf (last visited on 28 November 2014).

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational enquiry. *Educational Researcher, 32,* 9-13.

Dawes, R. M. (2001). Probabilistic thinking. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 12082-12089). Amsterdam: Elsevier.

Department for Education (2013) Mathematics programmes of study: key stages 1 and 2. National curriculum in England. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/335158/PRIMARY_national_curriculum_Mathematics_220714.pdf (last visited on 28 November 2014).

Fischbein, E. (1987). *Intuition in science and mathematics*. Dordrecht: Reidel.

Fischbein, E., & Gazit, A. (1984). Does the teaching of probability improve probabilistic intuitions? *Educational Studies in Mathematics, 15*, 1-24.

Gigerenzer, G. (2002). *Reckoning with risk*. London: Penguin Books.

Hart, K. (1981). Ratio and proportion. In K. Hart (Ed.), *Children's Understanding of Mathematics: 11-16* (pp. 88-101). London: John Murray.

Hart, K., Brown, M., Kerslake, D., Küchemann, D., & Ruddock, G. (1985). Chelsea Diagnostic Mathematics' Test: Teacher's Guide. Windsor: NFERNelson.

Lobato, J. (2003). How design experiments can inform a re-thinking of transfer and vice versa. *Educational Researcher, 32*, 17-20.

Murata, A. (2008). Mathematics teaching and learning as a mediating process: the case of tape diagrams. *Mathematical Thinking and Learning, 10*, 374-406.

Ng, S. F., & Lee, K. (2009). The model method: Singapore children's tool for representing and solving algebraic word problems. *Journal for Research in Mathematics Education, 40*, 282-313.

Nunes, T., Bryant, P., Evans, D., Gottardis, L., & Terlektsi, M.-E. (2014). The cognitive demands of understanding the sample space. *ZDM Mathematics education*, 46, 437-448.

Nunes, T. (1993). Learning mathematics: Perspectives from everyday life. In R.B. Davis & C.A. Maher (Eds.), *Schools, mathematics, and the world of reality* (pp. 61-78). Needham Heights (MA): Allyn and Bacon.

Nunes, T. & Bryant, P. (1996). *Children doing mathematics*. Oxford: Blackwell.

Nunes, T. & Bryant, P. (2009). *Key understandings in mathematics learning. Paper 4: Understanding relations and their graphical representation.* http://www.nuffieldfoundation.org/sites/default/files/P4.pdf (last visited on 28 November 2014).

Nunes, T., Bryant, P., Barros, R., & Sylva, K. (2012). The relative importance of two different mathematical abilities to mathematical achievement. *British Journal of Educational Psychology, 82*, 136–156.

Nunes, T., Bryant, P., Evans, D., Bell, D., Gardner, S., Gardner, A., and Carraher, J. (2007). The contribution of logical reasoning to the learning of mathematics in primary school. *British Journal of Developmental Psychology, 25*, 147-166.

Nunes, T., Bryant, P., Evans, D., Gottardis, L., & Terlektsi, M.-E. (2014). The cognitive demands of understanding the sample space. *ZDM Mathematics education*, 46, 437-448. DOI 10.1007/sl 1858-1014-05813.

Piaget, J., & Inhelder, B. (1975). *The Origin of the Idea of Chance in Children*. London: Routledge and Kegan Paul.

Schoenfeld, A. H. (1988). When good teaching leads to bad results: The disasters of "well-taught" mathematics courses. Educational Psychologist, 23, 145-166.

Streefland, L. (1984). Search for the roots of ratio: some thoughts on the long term learning process (Towards...a theory): Part I: Reflections on a Teaching Experiment. *Educational Studies in Mathematics*, *15*, 327-348.

Streefland, L. (1985). Search for the roots of ratio: some thoughts on the long term learning process (Towards...a theory): Part II: The Outline of the Long Term Learning Process. *Educational Studies in Mathematics, 16*, 75-94.

Verschaffel, L. (1994). Using retelling data to study elementary school children's representations and solutions of compare problems. *Journal for Research in Mathematics Education, 25*, 141-165.