Chinese Students' Understanding of Probability

Li Jun

School of Science

Nanyang Technological University

June 2000

Chinese Students' Understanding of Probability

Li Jun

School of Science

A thesis submitted to the

National Institute of Education,

Nanyang Technological University

In fulfillment of the requirement for the degree of

Doctor of Philosophy

June 2000

Abstract

In our modern society, people are faced more and more often with making decisions in an environment that involves uncertainty. Within this environment the teaching of probability is an important topic. The literature suggests that probability is a complex concept with many dimensions. Probability can be interpreted descriptively using words such as never, impossible, unlikely, probably, certain, and so on, but how they are used in probability may be different from the real-life use of these words. Probability can also be interpreted quantitatively using three approaches: theoretical, empirical and subjective. These approaches are complementary, since different approaches could be appropriate in different situations. However, it should be noted that in some situations more than one of the three approaches could be applied in the same situation.

Researchers who have investigated probability have identified many misconceptions, such as representativeness, availability, outcome approach, equiprobability, and so on. The results of the research show that the use of some misconceptions decreases with age, while others are very stable and even grow stronger with age. The research has usually been undertaken in Western countries.

This study investigated the following three questions: What are the main misconceptions of probability Chinese students have? What is the developmental structure of students’ understanding of probability? Can an activity-based short-term teaching programme improve ordinary grade 8 students' understanding of probability?

The research was divided into two parts. The first two questions were answered in the first part, referred to as the main study. The sample was 567 Chinese students from three grades (6, 8 and 12) and two school streams (ordinary and advanced). Eighty-three items, most in multiple-choice plus explanation form, in four categories (identification of impossible, possible and certain events; interpretation of chance values; chance comparison in one-stage experiments and chance comparison in two-stage experiments) were organised into nine distinct questionnaires. Sixty-four out of the 567 students were interviewed the day after the questionnaires had been administered.

The second part is referred to as the teaching intervention. Six activity-based lessons which focused on empirical probability were given to two grade 8 classes (each with about 25 students) in an ordinary school. The approaches were parallel except that one class had the opportunity to see computer simulations of a long series of experiments, while the other class was given the data in written form. During most of the teaching time the two classes did the same activities. All the students were tested and interviewed both prior to and after the teaching intervention.

Fourteen groups of misconceptions were observed in this study. The outcome approach, chance cannot be measured mathematically, compound approach and equiprobability were the main misconceptions for each grade and each stream of students. The context and data used in an item were found to play a role in eliciting some misconceptions.

The SOLO taxonomy was used in this study to describe students' hierarchical understanding levels on the concept of probability. It was found that, generally, there was no improvement in developmental level at grades 6 and 8, the two grades without any formal probability training. Grade 12 students have a better understanding than the younger students.

It was found that students' understanding of the frequentist concept of probability was the weakest. Most students in this study applied at least one misconception related to the frequentist definition of probability in their written questionnaire.

The results of this activity-based short-term teaching programme show that even a short intervention can help students overcome some of their misconceptions, such as chance cannot be measured mathematically. However, in this particular teaching experiment there was little change in the students use of the outcome approach and equiprobability, but it is possible that an alternative teaching experiment designed specifically to overcome these misconceptions might have a positive impact. Students in the two classes, one class with and one class without computer simulations, improved substantially in their answers and reasoning but no statistically significant difference was found between the classes.

Probability is not part of the present Chinese school curriculum, except in a very few cities such as Shanghai where there is a very limited unit in grade 12 of approximately 8 hours. This situation is currently under review and the findings of the study can be used to inform change. For example, the results show that Chinese students develop misconceptions about probability prior to any formal introduction. In introducing probability, this information needs to be considered, and the data from the teaching intervention shows that an activity-based teaching programme, whose design considers the specific misconceptions that students have, can be effective in overcoming some misconceptions, even when computer simulations are unavailable.

Acknowledgements

From June 1997 to June 2000, I studied at Nanyang Technological University as a PhD student. During these three years I learned a great deal from many people. Without their help or advice, I could not have completed the thesis.

Before I started my studies in Singapore, I only knew a little about the teaching and learning of probability. Through the literature I became acquainted with the research of Clifford Konold, Jane Watson, and Marie-Paule Lecoutre, and when I contacted them they sent many papers that helped me focus my research. Further, Clifford Konold, Jane Watson, Jonathan Moritz, Joan Garfield, and Carmen Batanero read some preliminary materials including test items or the criteria for data coding and their suggestions were very helpful for this study.

I also want to express my thanks to all the Chinese students involved in this study. Although they had no prior or very limited prior experience with probability, most of them gave quite detailed and serious responses to the questionnaires and interviews. Also, since the teaching intervention was undertaken outside school time, I appreciate the additional effort this group of students gave me.

My thanks also go to my colleagues in East China Normal University who helped organise the collection of data in Shanghai.

My supervisor, Lionel Pereira-Mendoza gave me great support, encouragement and advice. Most importantly, he taught me how to do research and how to report the research. I really appreciate the three years under his supervision that I believe provides me a sound basis on which to build my academic career in the future.

Much gratitude is due to other staff at the National Institute of Education (NIE), Lee Peng Yee, Luo Guanzhong, Phang Lay Ping, Yap Sook Fwe, Fan Lianghuo and Cheng Yuanshan. They gave me academic support and advice. I also like to thank two children of Lim-Teo Suat Khoh who allowed me to observe them being interviewed on their understanding of probability.

I would like to acknowledge the financial support for data collecting both from the School of Science of NIE and Southeast Asia Mathematical Society.

My schoolmates in NIE gave me support and provided a friendly, happy environment for me when I away from home.

Finally, thanks to my parents, my husband and my son. Without their love and support, I could not have left them to study alone in another country.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

Decision-making involving uncertainty is an integral part of people's life. Football fans try to win money in football pools. Weather forecasters use "the probability of rain" in their reports. Doctors recommend an operation but might say that there is a 90% chance of it being successful. Other words such as, unlikely, possible, probable, fair, possibility and so on, are also used in probabilistic situations, where one cannot completely determine the outcome in advance.

Since a lot information around us is expressed in probabilistic terms and the theory of probability can help us to understand uncertainty better, it is reasonable to include probability in school curriculum. Actually, "there is a growing movement to introduce elements of statistics and probability into both the secondary and even the elementary curriculum, as part of basic literacy in mathematics" (Garfield & Ahlgren, 1988, p. 44), a position supported in a more recent paper by Nemetz (1997).

The question arises as to what specifically should be taught, when and how. National curricular reform documents in countries such as the United States (NCTM, 1989), Australia (Australian Education Council, 1991) and New Zealand (Ministry of Education, 1992) provide information on the decisions made in each of these countries. They all recommend that students should study probability from early in their schooling, and probability should be taught within the context of interpreting data that is collected by students. For example, the *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989) advocates, "in grades K-4, the mathematics curriculum should include experiences with data analysis and probability so that students can … explore concepts of chance" (p. 54). In grades 5-8 and grades 9-12, the NCTM recommends extending the students' experiences with simulations and experimental probability to further refine their understanding. In addition, teaching resources, activities, programmes and software such as C*hance and Data Investigations* (Lovitt & Lowe, 1993), *Interactive Mathematics Program* (1998) and *Resampling* (Simon, 1992) have been developed and are available to support the teaching of probability.

However, one practical problem associated with the teaching of probability is the preparation of teachers. Watson (1995) stated that "many high school teachers find that they are not adequately prepared by their own education to teach these topics" (p. 121). Teachers are inadequately prepared both in their knowledge of probability, per se, and pedagogy. Probability courses usually appeared in tertiary level after the 1970s and the approach was quite formal, although in the 1990s computer simulations were introduced in some courses. Such an approach should not be transferred directly into elementary or secondary school level. In addition, since probability has only recently been introduced into school, there is a limited but expanding knowledge on how to help students learn probability.

This introduction will now examine the questions of why research into students' understanding is useful, why probability was chosen as the topic to study and why data was collected in China. Then, a perspective on China's school system will be provided. Finally, the chapter will conclude with the research questions and the significance of this study.

Students' Understanding --- A starting point for teaching

According to Freudenthal (1973), the best way of teaching a topic is to establish strong relationships both *inside* and *outside* of mathematics; that is, relate what will be taught to what has been taught (Hawkins & Kapadia, 1984, p. 362, called it internal connectedness) and to what has been experienced in the real world (external connectedness).

A typical example of an approach that ignored the outside relationship and over emphasised abstraction was the "New Maths" movement of the 1960s. The result was that the mathematics learned in school could not be applied in the real world and, consequently, was not retained or valued by the vast majority of students. A school curriculum that is developed from a mathematician's perspective, per se, and neglects the relationship between mathematics and students' reality will be ineffective.

In order to enhance the outside relationship, we have to seek answers to many questions, such as: What intuitive knowledge do students possess prior to introducing a topic? How is intuitive knowledge developed? What are the developmental frameworks that reflect students' understanding? What are the effective teaching strategies that can be used to overcome students' misconceptions and improve their understanding? Such questions are not new, but they have started to receive more attention.

Students' Understanding of Probability --- A difficult but potentially productive area for research

Compared to other mathematical topics, probability moved from a mathematical topic to become an integral part of the curriculum very quickly. Given that most of the research on students' probabilistic thinking appeared after the 1970s it is not surprising that our knowledge is still very limited. After a psychological and pedagogical review of children's conceptions of probability, Hawkins and Kapadia (1984) raised some key questions, such as: What conceptualisation of probability do children of various ages have? How might their conceptualisation be changed? What is the relationship between intuitive concepts and formal concepts, and so on. They concluded that, "In general, the present authors feel that these question have not yet been answered by the available research findings" (p. 374). Although in the past few years more research has been done, many problems are still unsolved, such as assessment tools, students' and teachers' conceptions and misconceptions, different culture influences, practical strategies for teaching and the role of metacognition in solving probabilistic problems (Shaughnessy, 1992).

Besides the short history of research on probability in schools, there are several things that make the research on students' understanding of probability complicated. First, the uniqueness of probability relative to other mathematics topics creates specific problems. In algebra, one can conclude that if $a > b$ and $c > d$, then $a+c > b+d$. Consider the following situation, which appears on the surface to parallel the above argument. For spinner A, its arrowhead has more chance of stopping in its red part. For spinner B, its arrowhead has more chance of stopping in its red part. Generally, it is not possible to conclude that the two arrowheads have more chance of both stopping in the red parts. Probabilistic intuitions are not always reliable. Formal probability theory can be counterintuitive.

Second, it is impossible to "prove" a theoretical probability by a trial or even a few trials. This conflict does not apply to other components of school mathematics. Simulations are useful in helping students understand the uncertainly and unpredictability of a single event, but an extensive set of simulations is needed to validate a theoretical probability. This lack of students' experience with experiments involving a large number of simulations hinders the development of probabilistic thinking. This situation is further complicated by the fact that different sets of simulations can produce different results. Furthermore, due to the irreversible nature of random events even concrete materials do not provide as valuable a support for checking results in probability to they do in other areas such as geometry or arithmetic.

Third, it is difficult to determine the actual nature of students' probabilistic reasoning. For example, students may obtain the same answer to a chance comparison task but use very different approaches, some of which could involve faulty reasoning. This means that in order to investigate students' thinking, we need to ask them to explain their thinking. However, because of the students' limited experience with probabilistic language, they often have difficulty in adequately explaining their thinking. This makes determining the students' rationale for their misconceptions complex. A good understanding of students thinking is essential for effective teaching.

All these reasons mean that investigating students' understanding of probability is a complex task but also a potentially productive area for research.

Chinese Students' Understanding of Probability -- An untouched research topic

Students in China have very limited experience with probability. There are many reasons for this due to influences from both inside and outside school. First, the role of statistics and probability in decision making is still underestimated. For many people, only a mathematician or a gambler needs probability. Also, there is little use of probabilistic arguments and language in the media. The second reason is the shortage of qualified teachers and teaching resources. Teachers have had very little or no exposure to probability. Even if they have received some exposure to probability in either pre-service or in-service training, the programme is usually dominated by theory, with no experience with activities or simulations. Furthermore, there are very limited teaching resources, such as textbooks, activities and materials (dice, marbles…) provided by schools. Third, probability is an optional topic in most areas of China. Only in Shanghai, is there an introduction of probability within the school curriculum. However, the major reason is the fact that probability is virtually non-existent in University Entrance Examinations. In China, the National University Entrance Examination for school teachers is described as the baton to a band. With the significance placed on these examinations, it is not surprising that research in students' understanding of probability has not been undertaken in China.

School System in China

Before proceeding further with the thesis, it is necessary to provide an overview of the Chinese education system.

Table 1.1 is a simplified framework for the system. For the past twenty years, in order to accelerate the training of talent, junior and senior high schools had been classified into two streams, advanced (containing about the top 10% of students) and ordinary. But the classification meant that competition among students, teachers and schools started as early as the first year of schooling and it led education in China in a wrong direction. So from 1997 the classification at junior level was eliminated and new students usually enrol in a junior high school near their homes.

Table 1.1 The Chinese Education System

Age	Grade	Structure	Notes
24 23 22		Postgraduate	
21 20 19 18		Universities, Colleges, Polytechnics	
		University Entrance Examination	End of secondary education
17 16 15	G12 G11 G10	Advanced Senior High Schools, Ordinary Senior High Schools, Vocational / Technological Schools	
		Streaming Examination (for Senior High School or Vocational School)	End of compulsory education
14 13 12 11	G9 G8 G7 G6	Junior High schools	
10 9 8 7 6	G5 G4 G3 G2 G1	Primary schools	
5 4 3		Kindergartens, Nursery Schools	

More than fifty years ago a few Chinese mathematicians tried to introduce statistics and probability into primary and secondary schools. However, in practice nothing happened until about 20 years ago when there was very limited introduction of statistics (Zhang, Zhou & Zhao, 1998). From 1978, descriptive statistics was included in the National Curriculum Standards, but it is taught only in one grade (grade 9) for about 9 total hours. The situation for probability is that it is still an optional topic in most areas of China. Shanghai was allowed to have its own curriculum standards in 1988.

The Shanghai curriculum standard included an introduction of probability and suggested about 8 total hours for the topic. It was covered in grade 12 after studying complicated calculations involving permutation and combination. The main focus is calculating probability, usually combined with permutation and combination calculations. There is no item on probability in National University Entrance Examinations since it still is an optional not a required topic. In Shanghai, in each of the 1998 and 1999 University Entrance Examinations only four out of 150 marks were allotted to probability. Here are the two items used. Students were asked to fill in the blanks.

> There are 4 white marbles and 3 black marbles in a bag. Pull out 3 marbles randomly. The probability that you only get one black marble is ______. (From the 1998 Shanghai University Entrance Examination)

> Roll a die twice and take the two numbers m, n as co-ordinates of point P. Then the probability that the point P locates within the circle $x^2 + y^2 = 16$ is ______. (From the 1999 Shanghai University Entrance Examination)

Although probability is still rarely taught at the school level in China, the situation is changing. A new National Curriculum Standard, which introduces chance and data into primary school level, will be issued in near future.

Research Questions of This Study

The main topic for this study is Chinese students' understanding of the concept of probability. Specifically, it investigated three main research questions:

1. What are the main misconceptions of Chinese students when answering chance interpretation and chance comparison problems?

2. What is the developmental structure of students' understanding of probability?
3. Can an activity-based short-term teaching programme improve Grade 8 students' understanding of probability?

Embedded in these questions are three components. The first component of this study, which parallels previous research, provides data on Chinese students' conceptualisation of probability. The second component, which extends previous research, explores students' misconceptions of frequentist probability and misconceptions associated with two-stage experiments that have only been briefly discussed in previous research. The last component is an investigation of the effect of an activity-based short-term teaching intervention based on the results of the first two components. It is designed to see whether such an intervention can both overcome some of the misconceptions identified in the study and improve students' understanding level of probability.

Significance of This Study

Although the importance of teaching probability is still underestimated in China, a new National Curriculum Standard, which plans to introduce probability into primary school and continually expand the teaching to secondary level, is being developed. Since this research studied school students' (of different ages and ability) understanding of probability the results will provide implications for the curriculum reform currently being undertaken in China.

Research on probability has mainly been undertaken in the West. Students in China come from a different culture and have different experiences with probability, both within the school setting and in everyday life. This study will extend educators' knowledge to students who have grown up under an eastern culture. As Shaughnessy (1992) said:

> Most of the psychological research on decision-making under uncertainty has been done in a very few countries, principally, in the United States, Israel, the United Kingdom, and Germany. What are the influences of culture on concepts of probability and statistics? Are phenomena like judgmental heuristics and misconceptions of probability just artefacts of western culture, or do they appear across many cultures? It would be interesting to see if misconceptions of stochastic and probability estimates under uncertainty vary across cultures. (p. 489)

This study involved three major dimensions, namely students' misconceptions, students' level of understanding and the effect of an activity-based short-term teaching intervention on students' misconceptions and level of understanding. Research on these dimensions should add considerably to the literature. It will provide information on Chinese students' misconceptions and understanding of probability. In addition, an activity-based approach to teaching the frequentist definition of probability to ordinary grade 8 students has rarely been done in China. The extensive nature of the study will help contribute to an overall view of students' understanding of probability.

Summary

In a modern society the ability to make judgements under uncertainty is becoming a part of basic literacy. Many countries have included, or at least recommended, that probability be an integral part of the school curriculum. An analysis of students’ misconceptions and developmental structure in understanding of probability is helpful in both developing appropriate curriculum and teaching strategies. Therefore, a study of understanding of probability for students of different ages and abilities in China is both timely in terms of current reform and will add to current knowledge of students’ probabilistic thinking. The next chapter will provide a review of the relevant literature.

Chapter 2 Literature Review

An examination of the literature on students' understanding of probability shows that this is a potentially productive area for research. This review focuses on three areas of probability research: students' misconceptions, probabilistic thinking framework and practical strategies for developing probabilistic concepts.

Research on Students' Misconceptions

In a modern society, people are faced more and more often with making decisions in an environment that involves uncertainty. The ability to make a good judgement in uncertain situations or to understand peoples' explanations of randomness is a key requirement to functioning effectively. It follows that probability needs to be an integral part of the school curriculum.

> Nevertheless, probability is a difficult subject to learn and teach. Causality is much more comfortable, logical thinking is much clearer, but chance is a reality. …In probability, paradoxes or counterintuitive ideas occur at the very heart of the subject, in the definition, and subsequently in relatively simple applications. (Kapadia & Borovcnik, 1991, p. 2)

As Kapadia and Borovcnik implied, the nature of probability suggests that teaching and learning probability is not an easy task. Students' misconceptions and difficulties in applying probabilistic notions require special study, especially if students have not been exposed to probability until the tertiary level or the later years of secondary level education, and they are familiar with logical or causal reasoning as the approach to mathematical thinking. In fact, as the subsequent review will show, many misconceptions in chance interpretation or chance comparison derive from the misapplication of definitive thinking.

In this study, a misconception in probability is intuitive knowledge that conflicts with formal probability theory. Both research on misconceptions in chance interpretation and in chance comparison are now reviewed.

Misconceptions in Chance Interpretation

The first skill that is required in discussing chance is to interpret chance qualitatively. Students are asked to identify possible, impossible and certain events and to use different words to describe the likelihood. After this experience, the idea of how to measure chance mathematically and interpret a chance value is introduced. Consequently, this part of the review is split into two components, interpreting chance qualitatively and quantitatively.

Interpreting Chance Qualitatively

There is some research that investigated, as part of their studies, students' qualitative interpretation of likelihood (Green, 1982; Fischbein & Gazit, 1984; Fischbein, Nello & Marino, 1991; Williams & Amir, 1995; Moritz, Watson & Pereira-Mendoza, 1996; Chan, 1997). Researchers employed two different types of tasks, one where the context was specified (for example, rolling a die once) and the other where it was not (for example, giving examples of impossible, possible or certain events). Although the research questions, the students' ages and their mathematical backgrounds varied among the studies, three general conclusions can be drawn. They are:

- Many students who have or have not received formal instruction in probability are able to identify possible, impossible and certain events and the ability improves with age.
- Confusion between rare and impossible, highly frequent and certain and confusion between certain events and possible events are two common misconceptions.
- In addition to the misconceptions, inadequate language ability is also a major cause of errors.

Fischbein, Nello and Marino (1991) undertook an investigation of 618 Italian pupils aged 9 to 14 years. The students were asked to identify events in given contexts then justify their answers. They found that the majority of students were able to identify the three kinds of events and their ability improved slightly with age. However, errors and misconceptions were not rare. Of the three kinds of events, certain events were the most difficult for the students (even for those who had received elementary instruction in probability). The researchers' explanation of the confusion between certain and possible events was, "usually, one tends to relate the notion of 'certain' to that of 'uniqueness'" (p. 527). Therefore, if a certain event is associated with a multiplicity of possible outcomes, "the notion of *possible* comes naturally into mind" (p. 527) and it is labelled wrongly as a possible event. They gave a few examples of students' justifications to support this explanation. In addition to this the researchers also mentioned other misconceptions, for example, confusion between uncertain and impossible and between mathematical meanings and subjective expectations, but these were not discussed in detail.

They further noted that the concepts of possible, impossible and certain are psychologically complex.

Other researchers (Green, 1982; Fischbein & Gazit, 1984; Williams & Amir, 1995; Moritz, Watson & Pereira-Mendoza, 1996; and Chan, 1997) used more difficult items. Students were required to **give examples** of impossible, possible/chance or certain events or **make sentences** involving these terms. The improvement with age for all three types of events and the confusion between certain events and possible events were observed again, but the results of an intervention in overcoming the confusion were different between Fischbein and Gazit (1984) and Chan's (1997) studies.

Fischbein and Gazit (1984) administered questionnaires to experimental groups (285 students) and control groups (305 students) of 10 - 13 year old students. The experimental group received instruction in probability. Questionnaire A, which was only administered to the experimental group, included some items involving possible, impossible and certain events. The data on these items show that almost all the 7th graders were able to give correct examples for possible and impossible events and more than 90% of the 7th graders were able to do so for certain events. The percentage of correct answers was much higher than in other research. This meant that although certain events were slightly more difficult than the other two events for the experimental group, the difficulty could be overcome by effective teaching.

Chan (1997) administered questionnaires to 425 students (16 -18 year olds) who had previously received instruction in probability. He also interviewed a subset of the students. He used an adapted but easier item where students were only required to give one example, while the original item required two examples (Original item: Fischbein & Gazit, 1984, Questionnaire A, the 8th item, p. 5), and added an explanation that a "certain event is an event that will definitely happen". His data show that even with this additional explanation more than 35% of the students gave wrong answers. In the majority of the wrong answers students gave a possible event as the example. He expressed surprise at the result since "numerous examples of different events were shown to the students in the lectures" (p. 53). Based on the data collected in the questionnaire and interviews he concluded that students might have two misconceptions. One is that a student might think that the term certain event meant the same as possible event. A student explained that both types of events had the nature that they were going to happen. The other misconception is that a student might think that each event consisted of exactly one outcome but she or he could not find any outcome that would definitely happen when rolling a fair die.

The data of Fischbein and Gazit (1984) and Chan (1997) were both collected after instruction but the percentage of correct answers was very different for the two studies, 98% versus 63% respectively. One possible explanation for the results is the different strategies used in the teaching. The specific examples used in teaching can play a critical role. As Fischbein, et al (1991) indicated trivial examples were not much help. They further stated that children "have to estimate the likelihood of an event by considering the joint outcomes which compose the respective event, and to avoid the spontaneous tendency to consider separately each outcome" (ibid. p. 528-529).

The literature reviewed to this point gives three explanations for the confusion between certain events and possible events: (1) relating the multiplicity of the outcomes to possible events; (2) each (simple) event consists of exactly one outcome; and (3) a certain event and a possible event are the same as both of them reflect that something is going to happen.

Green (1982) analysed the responses of 2930 students aged 11 - 16 (years 1 – 5 of secondary school) to a questionnaire on probability concepts. This questionnaire included items where students had to choose or give a word or phrase which means the same as terms such as impossible, possible, even chance and very probable. They were also asked to complete sentences in their own words such as "it is very likely that the Queen…". Green concluded that "pupils' verbal ability is often inadequate for accurately describing probabilistic situations" (p. 13). His data showed that the most difficult terms for the students were very likely, unlikely and very probable. Not very likely, little chance and by chance were easier and the easiest terms were likely, possible and impossible. Students had difficulty in giving a word or phrase which described the meaning of even chance, but they felt that choosing phrases that meant “has a 50-50 chance of happening” was much easier. He did not include an item that asked them the meaning of certain. The main errors reported were equating very likely to certain or always happen, and equating not very likely and unlikely to impossible or cannot happen. These findings are generally consistent with those of Fischbein, Nello and Marino (1991) and Williams and Amir (1995). Williams and Amir (1995) reported when they asked the subjects (11 - 12 years old) to quantitatively interpret how close the term was to impossible (0) or certain (6), most wrong answers were for the terms sure to happen and always happen and the best results were for the terms probable and even chance. Their results supported again the conclusion that students confuse rare and impossible, highly frequent and certain when answering items involving impossible, possible and certain events. It also showed how the nature of the tasks could lead to different conclusions. It is probably easier for students to quantify even chance on an interval between impossible and certain (the Williams and Amir task) than to explain even chance in their own words (the Green task). Williams and Amir also reported that many pupils did not answer the question involving the term seldom happens, and they suggested it was probably because the students did not know the meaning of the word seldom. The studies by Green (1982) and Williams and Amir (1995) highlight the linguistic difficulties associated with the terminology used in probability situations.

What does chance mean to younger students? What examples do they use to explain chance and in what contexts are these examples given? These are interesting questions and important for teaching. Williams and Amir (1995) found that children use the term chance in a number of different ways. For example, to describe an unplanned, unintended or unpredictable thing; to mean something has an opportunity or possibility of happening; to provide a measure of the likelihood of an outcome or event and to indicate taking a risk. They also found children's understanding of chance overlaps with their understanding of another word, luck, and is influenced by both religious and superstitious beliefs. Moritz, Watson and Pereira-Mendoza (1996) also investigated students' written responses and classified them into three broad categories: school-based, personal, and general/wider world. When they analysed the responses of students from Australia and Singapore they found, "students from Singapore were more likely to focus on events from the natural world, rather than consider human designed structures such as school schedules or games and competitions" (p. 21-22). Although they did not identify clear reasons for this, they highlighted the importance of context. Discussing probabilistic concepts in context permits students to consider multiple embodiments of a concept as well as confront their misconceptions. It makes the teaching more effective and meaningful to students.

This brief summary indicates that the qualitative interpretation of chance is a complex process. Students' interpretation of the meaning of these terms is a function of many variables such as the way the problem is posed and the context of the problem. Although many students are able to identify and give examples of possible, impossible and certain events or explain the different words when describing likelihood, there are a significant number of students who exhibit misconceptions. The review will continue with some of the research on the quantitative interpretation of chance.

Interpreting Chance Quantitatively

Interpreting chance quantitatively is needed to identify the likelihood of different events in a more specific manner. Before continuing further, it is necessary to explain the meaning of the term probability. There are three methods for assigning a probability to an event: namely, the classical, frequentist, and subjective definitions of probability (Konold, 1991; Johnson, 1992). The first, the classical interpretation, or theoretical interpretation, defines the probability of an event as the ratio of the number of alternatives favourable to that event to the total number of equally likely alternatives. It is an a priori assignment of probability. In other words, it allows for the calculation of probability based on a theoretical analysis, without the need for experimentation. The second, the frequentist interpretation, or empirical interpretation, defines probability in terms of the relative frequency of occurrence of an event in an infinite, or near infinite, number of trials. Therefore, it is an a posterior and experimental approach based on information after actual trials have been completed. The last one, the subjective interpretation, or intuitive interpretation, is a personal evaluation of the likelihood of chance phenomena. The initial estimation is subject to adjustment according to new information such as the results of actual trials. From a pedagogical view, all three should be taught during a student’s school career. The classical model is well defined and allows for an exact value of probability to be derived in situations where each outcome in the sample space is equally likely. The frequentist model has no such limitation and can be applied in situations where there is no theoretical way to derive a sample space of equally likely events. Subjective estimation is a good starting point for teaching since it can link the students’ own experiences with the theory of probability. Some of the literature on students' understanding of these three approaches is reviewed.

Theoretical Approach – the equiprobability bias and other misconceptions

Based on their experience of games involving activities such as rolling a die, tossing a coin, or drawing a slip of paper from a jar in a lucky draw, children easily develop the idea of equiprobability. Equiprobability means that each of the possible outcomes of an experiment has an equal probability of occurring. However, equiprobability is not valid in all probabilistic situations. The equiprobability bias refers to the misapplication of equiprobability in situations where the outcomes are not equally likely. Only under the condition that all the outcomes are equiprobable can we predict the probability theoretically by the classical approach.

The equiprobability bias is one of the main misconceptions in probabilistic thinking and has been the subject of considerable research. It is a very resistant bias (Monks, 1985; Lecoutre, Durand & Cordier, 1990) and is observed in students of various ages and mathematical backgrounds (Green, 1982; Monks, 1985; Lecoutre, Durand & Cordier, 1990; Fischbein, Nello & Marino, 1991; Lecoutre, 1992; Williams & Amir, 1995; Fischbein & Schnarch, 1997; and Cañizares & Batanero, 1998).

Green (1982), Williams and Amir (1995), and Cañizares and Batanero (1998) studied how equiprobability was applied to simple events. The students involved in the three studies were from 10 -16 years old. A lucky draw item, in which the student has to compare the likelihood of choosing a boy versus a girl when a teacher does a raffle in a class having 13 boys and 16 girls, was used in all three studies. There was considerable evidence of this bias. An average of 42% of the students in the Green study, 47% of the students in the Williams and Amir study and about 25% in the Cañizares and Batanero study exhibited this type of reasoning. A common conclusion was that this bias decreased with age, but there were still about 20% of the older students who answered that it is equally likely to pick out a boy's as a girl's name. Based on their study of 398 students aged 10 - 14 years, Cañizares and Batanero (1998) concluded that students' mathematics background was not an important factor in reducing the bias. They wrote,

> We did not observe any differences in the proportion of equiprobability bias between pupils with low and medium levels of mathematical performance and only a slight difference between the latter and those with a high level of mathematical performance (p. 1447).

Williams and Amir (1995) found that if the question was changed to 15 girls and 5 boys equal chance responses decreased substantially to 11%.

Equiprobability responses to simple event problems decrease with age. However, when students, even at the tertiary level, are faced with compound event problems the bias resurfaces. Green also included some complicated questions (more than two possible outcomes in a trial) such as asking which of a series of alternatives would happen most often (2H and 10T, 5H and 7T, 6H and 6T, 7H and 5T, all have the same chance) when 12 coins were tossed simultaneously. He found that approximately 70% of the students in every year showed an equiprobability bias by selecting the response that all have the same chance. For the same item, Monks (1985) reported that the equiprobable response was observed much less in bright children. In his experiment with 25 bright 14 years old boys, he found a short-term teaching programme involving computer simulation and general class discussions had a limited effect on removing the misconception. After the teaching intervention, his students improved substantially in answering some problems, but some students still had the misconception.

Lecoutre and her colleagues have been working on equiprobability since the1980s. A *standard problem*, comparing the likelihood of getting a five and a six versus two sixes when simultaneously throwing two dice, and equivalent problems were administered to university students with different mathematical backgrounds. After reviewing their series of investigations, Lecoutre (1992) classified the students' justifications into five cognitive models. First, the correct model leads to correct responses. Second, the construction versus chance model, when the chance aspect of the problem is masked, leads to correct responses, but when the mask is removed, the equiprobable bias resurfaces. Third, the conditional model leads to equiprobable responses. Students take two-stage experiment as two one-stage experiments and consider one as a condition of another. For example, given that one 6 has been rolled, since 5 and 6 are equally likely to be rolled, the likelihood of getting a five and a six versus two sixes is equal. Fourth, the chance model also leads to equiprobability. It means that random events should be equiprobable by nature. The last, the numbers model, leads to the conclusion that obtaining a pair of triangle cards is more likely when drawing two cards from a box containing two triangle cards and one square card, as there are more triangle cards in the box. She also indicated that about 50% of the students used the chance model in all their experiments, except in one experiment that they had specifically designed to remove the chance model.

In the studies by Fischbein, Nello and Marino (1991) and Fischbein and Schnarch (1997) more than 60% of the students in each level (from aged 9 to college students) discounting their backgrounds in probability, applied the equiprobability bias in solving the *standard problem* and parallel problems. By analysing the wrong responses the researchers also concluded that the chance model was one of the bases for this misconception.

The research on the equiprobability bias shows that taking all possible outcomes as equiprobable is a common and stable misconception held by many students. Chance events should be equiprobable by nature is the main justification for equiprobability. The review will now focus on other misconceptions associated with the theoretical concept of probability.

Piaget and Inhelder (1975) observed how children quantified probability in lot-drawing tasks. They found that young children had a variety of ways to compare the likelihood of a specific outcome when drawing a counter from each of two jars, which contained either the same number of their favourite counters or the same number of total counters. Concrete operational children begin to know how to quantify probability, but have difficulty in relating their favourite outcome, other possible outcomes and the whole sample space. For example, they compared only favourable cases or only unfavourable cases, or they compared differences between favourable and unfavourable counters and so on. These strategies often did not work when, for example, dealing with more difficult tasks where both the number of their favourite counters and total number of counters were different. Piaget and Inhelder concluded that concrete operational children regularly appear to be comparing only the number of target items (numerator in the classical probability calculation formula) and ignoring the number of total items (denominator in the formula). Only formal operational children who were able to construct part-whole relationships, which are needed to measure probability theoretically, were able to solve the difficult tasks using a calculation involving fractions. This analysis could imply that students below the formal operational stage are unable to learn probability.

However, “there is much psychological theory (and supporting evidence) to suggest that Man may be capable of probabilistic judgements from the very earliest stages of cognitive development" (Hawkins & Kapadia, 1984, p. 353). For example, Acredolo et al. (1989) challenged Piaget and Inhelder's conclusion that young children only attend to the number of target items. They wrote, "all age groups [from 7- 11] attend to variations in the denominator as well as to variations in the numerator, and, furthermore, that they attend to the interaction between these variables" (p. 933). They thought that the choice paradigm (which jar offers the best chance of yielding a target item in a random draw) was not an appropriate task for testing a child's ability to quantify probability. They seemed to feel that Piaget and Inhelder’s conclusion could be a consequence of the task, per se. To determine the validity of this conjecture instead of using pen-paper multiple choice test items, they asked subjects to mark chance size on a slide. In front of the slide images of a happy and a sad face were placed at the left and right ends, corresponding to certain and impossible. On the reverse side facing the experimenter was a ruler with numbers. They analysed the variation of students' estimation of the probability when the researchers change only one variable corresponding to either the numerator or denominator in the classical formula. They concluded that children used both.

In addition to the qualitative items on impossible and certain discussed earlier in Fischbein and Gazit's research (1984), students in the experimental group were given items where they were required to calculate probabilities. They found that a systematic programme on probability might be carried out without particular difficulties, possibly starting from grade 6 and certainly starting from grade 7. Most grade 7 students in their experiment were able to solve probability calculation problems independent of whether the condition involved replacement. They observed three typical incorrect solutions.

> One was the tendency to relate the two sets of marbles involved, one to the other, instead of relating the set of expected outcomes to the whole set of possible outcomes. The second misconception was that of expressing the probability as a 1/n ratio (because one extracts in fact, at each trial, a single marble). A third category of errors was of the following type: in a problem in which an extracted marble is not replaced the child forgets that one has to diminish by 1 both the number of expected outcomes and the number of all possible outcomes. (p. 9)

The review will now consider the research on misconceptions associated with the empirical approach to probability.

Empirical Approach – the outcome approach and other misconceptions

Some students do not think that frequentist information obtained from a series of repetitions can be used for estimating the probability for a single trial. In other words, they do not have the idea of interpreting probability in a frequentist way. A main justification for them is the outcome approach reported by Konold (1989, 1991 and 1995). He described the outcome approach as follows:

> The salient features of this approach are (a) predicting outcomes of single trials, (b) interpreting probabilities as predictions and thus evaluating probabilities as either right or wrong after a single

occurrence, and (c) basing probability estimates on causal features rather than on distributional information. (Konold, 1989, p. 70)

For outcome-oriented students, "probability involves *deciding whether* an event will occur, rather than *quantifying how often* that event will occur" (Konold, 1991, p. 150). When they interpret the numerical probability of an outcome, they use a 50% chance as a guide to decide whether the outcome will or will not happen. If the chance is much higher than 50% and close to 100%, they think it will happen. If the chance value is much lower than 50% and close to 0%, they think it will not happen. In the case of the chance being exactly 50%, they may say "don't know" or "can't be decided. "

Previously, research was presented showing that students equate rare with impossible and highly frequent with certain (for example, Green, 1982; Fischbein et al., 1991). It would seem reasonable to conclude that in addition to making this error based on subjective judgements, at least some students make this error by using the outcome approach. For example, Fischbein et al. (1991) quoted a student who argued that it is impossible to get a 5 when rolling a die. The student said that it was, "because it is only one probability among 6" (p. 528).

Konold (1989) conducted two interviews with undergraduate students. He used seven problems that might result in an outcome-oriented student using that approach. In the weather problem, which is often referred to when talking about the outcome approach, he asked several questions related to the interpretation of a numerical probability. Students using the outcome approach explained, for example, the meaning of there is a 70% chance of rain as it will rain. When asked what they would conclude if it did not rain, they believed that the forecaster’s prediction would have been wrong or during the night, the precipitation or something changed because of other outside factors. They also judged that a forecaster's prediction was sub-optimal if it did rain, as the proposition should be given a higher numerical chance (Konold, 1991; Konold, 1989). For an irregularly shaped bone problem, he found that even when a summary of the results of 1000 trials was shown, some students still preferred to base their predictions on a visual inspection of the bone rather than on the available data. Students argued that properties of the bone are a more stable source of evidence when compared with frequency data, which can fluctuate from sample to sample. So there is no difference whether you roll 1000 or 2000 times, as "things change" (Konold, 1989, p. 80). The justification reminds us of the chance model that was mentioned in the discussion of equiprobability. Here, the chance model could be explained as nothing can be predicted for chance events (even across a large number of repetitive trials).

Williams and Amir (1995) also studied the outcome approach in a study of 11-12 year old children's informal knowledge of probability. They adapted Konold's weather problem and painted-die problem (see Konold, 1989) to a multiple-choice format. Their research results indicated that a considerably lower percentage of pupils used the outcome approach in the painted-die problem than in the weather problem. The possible reasons they suggested were, "the use of the 'outcome approach' is triggered more by context and wording than by individual orientation. Or perhaps children are more fickle, i.e., are more likely to be influenced by context than adults" (Williams & Amir, 1995, p. 22).

As mentioned previously when the difference between the number of boys' names and girls' names in a cap was increased, Williams and Amir found that the equiprobable response decreased considerably from 47% to 11%. They explained it as, outcome approach responses decreased as the difference increased. It should be noted that it can be very difficult to distinguish whether an equiprobable answer is based on the use of the equiprobability bias or the outcome approach, especially if the options *cannot compare*, or *I don't know* are not given or the probability value is not stated. For example, equiprobable answer with explanations, such, as I don't know, that's odd and the outcomes cannot be controlled are common (Piaget & Inhelder 1975, Borovcnik & Bentz 1991, Fischbein, et al. 1991 and Lecoutre 1992) and such responses could be the result of using either the equiprobability bias (chance model) or the outcome approach (a 50% chance means it can't be decided or don't know). When the option *cannot compare* or *I don't know* is added, there is still a problem, for example an answer could be the result of either insufficient knowledge of chance comparison or using the outcome approach (trying to predict the outcome of a single trial) or thinking it is completely impossible to measure probability. Therefore, in this author's research, the option that it is impossible to compare or impossible to measure were added in all chance comparison items.

The outcome approach is a common and stable misconception used by students of various ages. Now, the research on other misconceptions that related to interpreting probability empirically is reviewed.

Randomness is a basic idea in probability. It includes two aspects: (1) the uncertainty and unpredictability of the single event and (2) the patterns that emerge across a large number of repetitions of the event (Metz, 1997). Based on their real life experiences, students can accept the first aspect without much difficulty. However, it is not so easy for students to understand the second aspect, namely that repetition can help to find a pattern and the pattern obtained from a large number of repetitions is more reliable than that obtained from a small number of repetitions.

Piaget and Inhelder (1975) suggested after a child is clear that the outcome for a single trial is unpredictable, the child has to have appropriate experiences to overcome the misconception that the larger number of repetitions makes the outcome less regular. In their interviews with children they used a spinner with eight equal sectors and asked questions such as, "which is more probable: two on each color with sixteen spins, or about one hundred on each with eight hundred?" (p. 76). They concluded that children in the second stage of the development of the idea of chance understand the uncertainty and unpredictability of a single event but recognise the law of large numbers "only for relatively small numbers, and reverses itself when we go beyond about one hundred" (p. 79). At this level children believe it is more likely that each colour will be hit after twenty spins than ten spins. So for small numbers their intuition is in accord with the law of large numbers. However, when comparing a small number of trials with a large number of trials, for example 16 versus 1600 trials, the students think that two of each colour with 16 spins is more likely to occur than 200 of each colour with 1600 spins. Piaget and Inhelder provided two reasons for this extraordinary reversal. The first came from a failure to understand proportionality. "To grasp the law of large numbers, it is important first that one recognizes that a difference of five in one hundred for example is less a difference than five in ten" (p. 79). These children only compared the absolute values. The other came from the inability to handle large numbers. "For one hundred, eight hundred and sixteen hundred trials, he is no longer certain of what can happen because he lacks the ability to see the number of possible representations or directions" (p. 80). That is, students can only handle data at the concrete operational level, so they did not succeed in going beyond groups of one hundred trials. Piaget and Inhelder (1975) indicated only formal operational children answered with the understanding of the law of large numbers.

There is limited research that has focused on adolescents' understanding of frequentist probability. One study was by Lidster, Pereira-Mendoza, Watson and Collis (1995). They interviewed 12 Grade 6 students, 6 Grade 3 students and 6 Grade 9 students. Each interview lasted over 45 minutes with groups of three students. At the beginning of the interview, the interviewer played a simple dice game with the students and asked the students to record the outcomes. In the first two games, the interviewer took the loaded dice and the students had the fair one. Before the third game, the interviewer asked the students if they wanted to change dice. After several games, students were sequentially shown some prepared graphs. They represented the information from 60, 360, 600 and 12000 trials performed with six dice (some of the six dice were loaded). For each graph, the students were required to say whether the dice were fair, unfair or they were not sure, and what other information would they need to be certain of their decision. The last graph showed exactly equal outcomes for each number on a die and asked the students whether it represented the results of throwing a fair die. The students were then required to draw graphs of a fair die after 60 and 120 trials. Some of the students were also asked to draw a graph of an unfair die.

They found some students, especially younger students, answered according to their own intuitions and egotistic beliefs such as the way of rolling determines the result or no unfair dice exist in the world. But they also found that most students in their study "were willing to accept that a die was fair in the face of strong evidence that the die was most probably biased" (p. 17). They explained this in two ways. One was that some students did not know how an unfair die could be made so they believed all dice were fair. The other was their experience of playing games. When they lose a dice game they were told by their parents that the dice could come up with any result and it was still fair and acceptable. Misconceptions like neglecting sample size, comparing the absolute difference rather than the ratio of the number of outcomes in determining whether a die was fair were reported but without further discussions of their origins.

Subjective Approach – the representativeness and other misconceptions

What is the chance that a person will get a new job within two months? What is the likelihood that a company will win when bidding on a new project? There are a lot of chance problems in the real world that cannot be answered by either classical probability or frequentist probability. In such a situation, people always initially make a subjective or intuitive estimation based on the information available then adjust the estimation according to new information. However, sometimes such an adjustment leads to a worse estimation guided by misconceptions, such as representativeness.

A main misconception in quantifying chance subjectively is representativeness, summarised and studied by psychologists Kahneman and Tversky in the 1970s and 1980s. Tversky and Kahneman (1974) found that in a situation where an object or event belongs to another class or process, representativeness is usually employed.

> A person who follows this heuristic evaluates the probability of an uncertain event, or a sample, by the degree to which it is: (1) similar in essential properties to its parent population; and (2) reflects the salient features of the process by which it is generated. (Kahneman & Tversky, 1972, p. 33)

In other words, the more similar (representative) the situation is in terms of the properties of the parent population, the higher probability the situation has of occurring, and situations that are equally similar (representative) to their parent population have equal probability (Kahneman & Tversky, 1972).

The typical example of using representativeness is people believing that the sequence BGGBGB is more likely to occur in a family of six children than either BBBBGB or BBBGGG (Kahneman & Tversky, 1972, Shaughnessy, 1977). The sequence BBBBGB seems to include too many boys, and BBBGGG looks like too orderly. BGGBGB is the most representative so is the most likely one.

Consider the *standard problem* (comparing the likelihood of getting a five and a six versus two sixes, when simultaneously throwing two dice) that has been discussed earlier. Fischbein, Nello and Marino (1991) found that some students obtained the correct answer guided by the belief that "identical results appear less often than different results" (ibid. p. 535). They reported that Lecoutre and Durand had the same finding. Although the researchers did not mention representativeness, it seems an appropriate way to categorise the explanation.

During the past thirty years, mathematics educators (Shaughnessy, 1977; Green, 1982; Konold, 1989; Cox, & Mouw, 1992; Konold, Pollatsek, Well, Lohmeier & Lipson, 1993; Williams & Amir, 1995; Fischbein, & Schnarch, 1997) have reported evidence of representativeness and its related misconceptions. Based on data presented by these researchers, it can be concluded that students' responses are very dependent on the form of the task. When students were asked to compare <u>which sequence</u> is more likely to occur (for example, Kahneman & Tversky, 1972; Shaughnessy, 1977; and Konold, 1989), about half the students answered by using a representativeness approach. Alternatively, when students were asked to compare <u>which outcome</u> of the next trial is more likely to occur based on the given sequence (for example, Green, 1982; Konold et al, 1993; Williams & Amir, 1995; and Fischbein & Schnarch, 1997) the percentage of representativeness responses declined to less than 35% as many students used either an outcome approach or the equiprobability bias instead.

The dependence on the form of the task was supported by students' responses in a study by Konold et al (1993). They posed the four-heads problem in their interviews:

> A fair coin is flipped four times, each time landing with heads up. What is the most likely outcome if the coin is flipped a fifth time?
>
> a) Another heads is more likely than a tails
>
> b) A tails is more likely than another heads
>
> c) The outcome (heads and tails) are equally likely (p. 397)

They found that students wanted to know whether the problem referred to a single flip (the fifth flip) of the coin or whether it referred to a sample of five flips in which the first four flips are known to have come up heads. In the former situation (which outcome), students answered equally likely, but in the latter one (which sequence), they answered tails, which is consistent with a representativeness approach. This illustrates that students changed their thinking depending on how they interpreted the problem.

Konold et al (1993) also found a switch existed in another problem with two parallel parts.

The problem was the HT-sequence problem. In Part 1 of the problem the students were required to select which of the following five choices was most likely when a fair coin is flipped 5 times.

a) HHHTT

b) THHTH

c) THTTT

d) HTHTH

e) All four sequences are equally likely (p. 397)

In Part 2 they were asked which was the least likely when presented with the same a) - d) and e) was replaced with e) All four sequences are equally unlikely.

The correct answer to both parts should be e). They found that 72% of their subjects got the correct choice in the most likely part, but only 38% in the least likely part. They suggested that the majority of students used representativeness in the second part but not in the first part. They explained the high percentage of equally likely responses in the first part as the result of students using the outcome approach and called such an inconsistency the "M-L (most-least) switch".

However, this author suggests a possible alternative to the interpretation of the inconsistent results. The switch might not exist at all for some students as they applied representativeness consistently in the two parts. For part 1, the most likely part, almost the same number of students selected each of the four sequences as the most likely. The percentages for each sequence ranged from 2.3% to 9.3%, while in part 2, the least likely part, the range was far larger, from 2.5% to 29.1%. In other words, for the students none of the four sequences is more representative of the parent population, but some of the four sequences are less representative of the parent population than others. According to the representativeness approach, events that are perceived as equally representative of their parent population have an equal probability of occurring and one perceived as least representative of its parent population was the least likely. This led to the choice of equally likely in the most likely part but non-equally likely options in the least likely part. Therefore, it might be reasonable to explain that the high percentage of correct choices in the most likely part was the result of students' correct understanding, or their use of either an outcome approach or representativeness.

According to Kahneman and Tversky (1972) not only do students without prior training in probability rely on representativeness, but some psychologists who are familiar with data analysis use it unconsciously. Representativeness is a strong heuristic and difficult to overcome (Cox & Mouw, 1992). However, two teaching interventions undertaken by

Shaughnessy (1977) and Fischbein and Gazit (1984) showed that change was possible. Shaughnessy (1997) found that "mere exposure to probability concepts in a lecture format is not likely to be sufficient for overcoming the strong influence of the representativeness heuristic" (ibid. p. 310-311), but "experimental activity-based classes were more successful at overcoming reliance upon representativeness" (ibid. p. 308). Fischbein and Gazit (1984) found that both age and instruction had a clear positive effect on overcoming negative recency - a misconception related to representativeness. However, in a lottery sequence problem (whether a consecutive sequence has more chance to win than a more random appearing sequence) they concluded that "age itself does not solve the problem. It seems that only by way of instruction can this intuitive obstacle [representativeness] be overcome. This instruction effect is clearest at the 12-13 year old level" (p. 17).

The law of large numbers says that large samples are representative of the population from which they are drawn, but a parallel "the law of small numbers" is not valid (Kahneman & Tversky, 1972). Guided by the law of small numbers, people believe that samples are very similar to one another and to the population from which they are drawn. They also believe sampling is a self-correcting process. The two beliefs lead to some specific misconceptions. One is neglecting sample sizes and the others are negative recency and positive recency.

"Since the size of the sample does not reflect any property of the parent population, it does not affect representativeness" (Kahneman & Tversky, 1972, p. 38). So people may agree that the probability of obtaining an average height greater than 1.7m is the same for samples of 20, 100, and

200 men. Fischbein and Schnarch (1997) gave two problems that related to the effect of sample size to students in Grades 5, 7, 9 and 11 and prospective teachers. They concluded that "the basic misconceptions is that sample size is not relevant", and that "this misconception *developed* with age" (p. 101).

Believing sampling is a self-correcting process, in a two-outcome experiment many people believe that after a long run of one outcome the other outcome is more likely to occur. For example, when tossing a coin after a long run of heads a tail is more likely to come up for the next toss. This is called a negative recency effect. However, other people who believe even small samples must be representative of the parent population, conclude that the result of long run of one outcome makes it more likely it will continue to occur. For example, when tossing a coin after a long run of heads a head is more likely to occur for the next toss. This is called a positive recency effect. They draw opposite conclusions from the same information.

The representativeness heuristic was also applied in studies of randomness (Kahneman & Tversky, 1972, Falk, 1981; Green, 1982). These researchers found that sequences that maintain an approximately equal number of the various possible outcomes and alternate irregularly between possible outcomes are judged as more random than more regular situations.

In conclusion, each of the three approaches to probability has specific main misconceptions, i.e., equiprobability, the outcome approach and representativeness. It is a complex task to design teaching strategies that will help students to quantify probability. It is also a complex task to determine which misconception a student really used in his or her justification, as sometimes different misconceptions lead to the same conclusion. This study focused on the first two approaches to probability, theoretical and empirical approach. The review will now focus on chance comparison.

Misconceptions in Chance Comparison

Many misconceptions and strategies have been reported in comparing chance. This part of the review will involve two sections. The first is on one-stage experiments (such as drawing a single marble from a bag) and the second is on two-stage experiments (such as drawing two marbles, one from each of two bags). Since errors that are caused by the equiprobability bias, the outcome approach and representativeness have been discussed earlier, this component of the review will focus on other special methods used by students.

Chance Comparison in One-Stage Experiments

Besides the equiprobability bias discussed previously, students who have not received instruction in probability often use self-designed rules when they are required to compare likelihood. Piaget and Inhelder (1975), Noelting (1980a, 1980b), Falk, Falk, and Levin (1980), Falk (1983), Green (1982, 1983a, 1983b), Singer and Resnick (1992), and Watson, Collis, and Moritz (1997) studied these rules in their research. Although the students' ages in these studies ranged from 4 to16 years and more ikonic responses were observed in younger children, the tasks they gave the children were similar and all followed the basic structure of Piaget and Inhelder's classic lot-drawing tasks (Piaget & Inhelder, 1975, p. 133) involving two bags with a given number of coloured marbles. Students are asked to determine which of the bags gives a greater chance of picking the preferred coloured marble or whether the two bags offer the same chance. By changing the proportions of components in the two bags different strategies were identified. Green (1983b) summarised four common identifiable strategies used by his subjects (aged from 11 -16). They are: (1) choose the bag with more total counters (percentage of sample in his study is 1.4%); (2) choose the bag with more black - the colour wanted - counters (7.5%); (3) choose the bag with the greater black/white difference (3.1%); and (4) choose the bag with the greater black/white ratio (19.9%). Only 32% of the pupils consistently applied one pure strategy to all his five problems (same context, different proportions of components). Choosing the bag with more of the wanted coloured counters was the common pattern observed among younger pupils. He also found that the correct proportional reasoning strategy grows with age and the other three biases decline. Falk (1983) interviewed 35 children aged from 5 to 10 years old and reached the same conclusions as Green. Her statement on the design of the instruments is important. She found that some of the incorrect choice strategies were not and could not be discovered in some of the previous research because of the limitations of their design. This was the origin of some researchers' overestimation of children's ability to understand probability. Such a conclusion was also suggested by Acredolo et al. (1989). Falk (1983) suggested further,

> We should take care that this would not happen in the teaching of probability. The teacher's responsibility should be to present the concept in its most general form and to make sure that all the components of the concept (namely, the numbers of elements of both

categories, the difference between them, and the total) will appear in all their possible mutual relations and will vary in a free interplay (p. 725).

Chance Comparison in Two-Stage Experiments

In order to compare chances theoretically, we need to know how many possible outcomes will actually occur or the sample space of an experiment. This task becomes more difficult when solving a multi-stage problem as the number of sample points is often much larger than in single-stage problems. For example, when rolling one die there are 6 possible outcomes while when rolling two dice there are 36 possible outcomes. A situation with many outcomes provides a real challenge to students who have never been exposed to combinatorics problems. There were two difficulties. One was that students had no clear way to generate the sample space. They listed the outcomes in a trial-and-error way (Piaget & Inhelder 1975; Green 1982; Williams & Amir 1995; English, 1993) or could not give the correct number for all the possibilities (Fischbein, Nello & Marino 1991). For example, "Many subjects used the additive procedure for calculating the number of all possible outcomes, instead of the appropriate multiplicative procedure (i.e., 6+6=12 instead of 6·6=36)" (Fischbein & Gazit 1984, p. 21). Another difficulty was an unconscious perception of the order. For example, they did not perceive that the result of rolling a 5 and 6 can be obtained in two ways (Lecoutre, Durand & Cordier, 1990; Fischbein et al., 1991). However, with the help of concrete and familiar settings, such as the lining up of three children or the dressing of toy bears in all possible combinations of tops and pants, most pupils were able to give all possible outcomes in two-stage or three-stage experiments (Green, 1982; English, 1993). Given a hint in a task such as comparing the likelihood of obtaining the same number or different numbers with both dice, many older students displayed "a natural, intuitive tendency to evaluate the probability of a compound event on the basis of the corresponding sample space" (Fischbein, Nello & Marino 1991, p. 546). But far fewer pupils could proceed to a formal level item of listing sample points and only a few of pupils considered the order of the elementary results which constituted the outcomes (Piaget & Inhelder 1975; Green 1982; Fischbein, Nello & Marino 1991; Jones, Langrall, Thornton & Mogill, 1997). Fischbein et al. (1991) concluded, "There is no natural understanding of the fact that, in a sample space, possible outcomes should be distinguished and counted separately if the order of their elementary components is different" (p. 547), a conclusion supported by Fischbein and Gazit (1984) and Lecoutre (1992). Lecoutre and her colleagues have tried various methods to lead the students to notice the order (for instance, colouring the two dice differently), but this did not have a significant effect (Lecoutre, Durand & Cordier, 1990).

In addition to the difficulty with sample space, another main misconception is to split multi-stage tasks into several independent one-stage tasks (Fischbein, Nello & Marino 1991, Konold et al, 1993, Benson & Jones, 1999). When analysing results for the *standard problem* mentioned previously, Fischbein et al. (1991) found that fewer than 10% of the older pupils who had received instruction in probability got the correct answer that P(6,6) is not equal to P(5,6 and 6,5). The majority said that they had the same chance. One student explained his reason this way, "Each die is independent from the other. The probability that with one die, one will obtain a certain number is 1/6 and it is the same probability that one will obtain the same number with the other" (p. 535). The researchers pointed out that the student used a "more sophisticated idea that 5 and 6 are equiprobable and therefore every event [indicates the event "double six" and the event "one five one six"], representing a binary combination of them, has the same probability" (p. 535). The explanation of the conditional model (Lecoutre 1992) and an answer to the same problem cited from a study of Lecoutre and Durand may support the same thinking: "Same probability. The two results (5 or 6), have the same chance". (Fischbein et al., 1991, p. 533). Since this idea is of particular interest in this thesis it is defined as the **compound approach**. The compound approach is a misconception used in measuring and comparing probability in multi-stage experiments. It splits multi-stage experiments into several one-stage experiments and interprets the likelihood of a multi-stage event as just a combination of the likelihood of its elements.

Before continuing with the next section of the review, research on students' thinking framework, a brief summary of some of the key conclusions of this section of the literature is included. The results reported here show that understanding probability is a very complex task. Even the identification of possible, impossible and certain events is more difficult than might have been expected. The equiprobability bias, the outcome approach and representativeness are three common misconceptions that correspond to classical, frequentist and subjective probability notions. It seems the large number theorem associated with the frequentist approach is a difficult concept for students.

Even young children have some intuitive strategies in comparing chance in one-stage experiments, but these intuitive strategies are usually unworkable in multi-stage experiments. As they lack knowledge of sample spaces, they may use the compound approach to split the multi-stage experiment into several one-stage experiments, consider the likelihood of each elementary result separately, and then combine them.

It should be noted that based on a written explanation it is sometimes difficult to identify which misconception underlies the response. So carefully designed instruments requiring justifications and interviews are important in identifying students' misconceptions.

Research on Probabilistic Thinking Framework

Students' misconceptions are often related to their limited cognitive development in probability. Therefore, the investigation of students' developmental structure in the understanding of probability is important. In this part of the review, the theory of children's cognitive development in probability developed by Piaget and Inhelder and a neo-Piagetian theory SOLO taxonomy are summarised followed by a discussion of the application of SOLO in assessing students' responses.

The cognitive-developmental theory of Piaget and Inhelder

Piaget and Inhelder undertook the earliest and most comprehensive research on the developmental understanding of probability. In their book, *The Origin of the Idea of Chance in Children* (1951, translated into English in 1975), they presented details of their systematic clinical interviews with children. They described three major stages in children's development of probability.

According to their theory, in the first stage, pre-operational children (prior to about age 7 or 8) have no means of discriminating between caused and chance events. The child at this stage tries to find order within disorder, as he or she believes that hidden order exists. The child imagines that he or she can predict the outcome, either in each isolated case or as a function of the previous cases. The child believes an outcome that has not yet occurred has more chance of occurring than one that has already occurred as each one gets its turn. When the child "is not successful in predicting isolated fortuitous cases by basing his answers on compensation, he frequently begins basing his predictions on the largest frequencies observed up to that time" (Piaget & Inhelder, 1975, p. 220). The child's explanation is that "they come out more easily than the others". Faced with a set of coloured counters where the numbers of each colour could vary, the child is not concerned with the quantities but simply makes his or her decision "by an examination of the favourable cases without taking into account the possible ones" (p. 135). Actually, the researchers "do not yet find at this stage any reasoning related to the whole field of distributions (large numbers, etc.)" (p. 222). When faced with the situation that a special result occurs again and again, the child always accepts the result without surprise. The child explains the unusual result by phenomenal or egotistic reason, such as "that can happen," "it's chance," or "it is a trick you can do with your hand". These are the same explanations the child gives when explaining a usual result. When asked to give all possible combinations of two or three elements, the child is unable to list all possible outcomes and has no idea when the series is finished.

In the second stage, a concrete operational child, from ages 7 or 8 to about 12, is able to differentiate the possible and the necessary. The child begins to know how to quantify probabilities, but with an incomplete set of strategies for calculating odds in complex situations. When given experiments without replacement, the child does not realise the changing composition of the whole, and "thus he considers only the initial composition and not the variable proportions" (p. 228). The child refuses to predict the outcome in a single trial situation, saying that it could be any possibility, but the results of several successive trials allow the child to recognise a regular distribution. However, the ability to recognise this regularity does not increase for these subjects with very large numbers. On the contrary, the child believes that regularity can occur "more easily with fewer trials" (p. 79). When faced with the situation that a special result occurs again and again, the child at the second stage of development realises the difference between "fortuitous dispersion" and "causal regularity". The child tries to find a reason to explain why an extremely unlikely outcome happens so often. The child understands the possibility of predicting all the arrangements of 2, 3, even 4 elements. But his or her approach is empirically based and incomplete.

Finally, formal operational students, from about age 12, are able to integrate deductive logic and concepts of chance. A child at this level realises that after 1600 trials the distribution information will be more representative than after 16 trials. Faced with a set of coloured counters where the numbers of each colour could vary, the child can make precise calculations. When presented with a situation where a special result occurs again and again, the child is able "to reconstruct reality by a hypothetical-deductive method" (p. 93) and use a systematic method automatically to discover the reasons. When required to give all the combinations of two or three elements, the child is able to complete the work systematically and generalise the conclusion to the problems involving more elements.

Although Piaget and Inhelder's research provides an overview of what children know at various ages on the development of the idea of chance, several concerns have been raised regarding some of the conclusions. The first relates to potential problems arising from their research methods.

> Like most Piagetian research, the methodology requires a high degree of verbalization from the children, and the task selection and task administration can be accused of begging the research question it proposes to answer-- namely, that there are these levels.
>
> (Shaughnessy, 1992, p. 479)

By analysing some examples in Piaget and Inhelder (1975) it is possible to identify responses that may have been influenced by the questions the children were asked. For example, students are shown two sets of counters, each set having some counters with a cross on one side. The sets are then mixed up separately and the students are shown two sets (with the crosses hidden). They are asked "to decide in which of the two sets he has the greatest chance of finding a cross on the first try." (p. 132). Different combinations of sets are used in their interviews.

Chan was given a problem in which one set had one counter with a cross and one counter without a cross (1/2) and the other set had two counters with crosses and two counters without crosses (2/4). Chan, who was identified as being at the first stage, answered as follows.

Chan: *There* (1/2*), no, here* (2/4) *because there are two crosses. No, it's easier here* (1/2).

R: And this way (problem with 1/3 and 2/6)?

Chan: *Here* (1/3), *there's a better chance*. (p. 146)

Another child Kel, who was identified as being at the second stage, when asked the same question answered as follows.

Kel: *Here* (1/2) *because there are fewer. No, this one* (2/4*) because there are two crosses*.

R: Is it the same?

Kel: *No, this one* (1/2*) is more certain anyway…. No, both because here* (2/4*) there are two of each, and here* (1/2)*, one of each.*

R: Very good. Now explain it. … (p. 154)

The question can be asked as to whether the explanations and the resulting categorisation of the students are a function of the questions. What would Chan have said if he had been asked the question, "Is it the same?" It is possible that his response would have led the interviewee in a different direction. According to Kel's answers to other items such as a problem with 1/3 and 2/6, it looks as if his equiprobable response was fragile. It is also possible that if he had not been asked the question he might be categorised at a lower development level. Therefore, there seems to be some concern as to whether the problems and questions posed lead to an accurate reflection of the child's developmental level.

The second concern is the relationship between age and the stage model of development. Acredolo, O'Connor, Banks, and Horobin (1989), said that first- through fifth-grade children "attend to variations in the denominator as well as to variations in the numerator, and, furthermore, that they attend to the interaction between these variables” (ibid. p. 933). This finding contrasts sharply with Piaget and Inhelder's (1975) contention that for concrete operational children "the comparison of two proportional groups, or groups where the whole and the part vary simultaneously, still remains impossible" (ibid. p. 230). It appears that younger children can understand probability.

The final concern is with the teaching implications of the research. “Piaget has also been criticized for putting too much emphasis on the child's failures, rather than on success” (Smith & Cowie, 1988, p. 300). The Piagetian perspective is that it is the child rather than the teacher who decides when to start the teaching. However, according to Fischbein (1975), instructional intervention at school plays an important role in developing the child's idea of probability.

> We have demonstrated that, through the use of figural procedures, schemas considered by Piaget and Inhelder to be accessible only at the level of formal operations can be constructed at the level of concrete operations. At the least, we have shown that the absence of proportionality is not an obstacle to learning the concept of probability. Even before the age 10, the child is capable of assimilating this schema with the help of elementary instruction. (Fischbein, 1975, p. 123)

The framework developed by Piaget and Inhelder had been the basis for later frameworks for describing students thinking. The framework that will be used in this study is the SOLO taxonomy and this is now briefly described.

The SOLO Taxonomy

Starting in the 1980s, theories of cognitive development extended their interests beyond logical operations as well as putting an increased focus on the relationship between their theories and teaching practice. "These new directions have influenced recent views on school curriculum planning, assessment and, in consequence, school learning" (Collis, 1992, p. 25). Biggs and Collis (1982) have a different perspective from that of Piaget. They believe the development and structure of learning are most appropriately discussed in terms of student responses and not a student, per se. They developed a neo-Piagetian model -- the SOLO (**S**tructure of the **O**bserved **L**earning **O**utcome) taxonomy (Biggs & Collis, 1982; Biggs & Collis, 1991 and Collis & Biggs, 1991). Like some other neo-Piagetian theorists they use increasingly sophisticated function modes and an increasingly complex structure of responses within each mode to determine the level of an individual's response (Collis, 1992). There are two main differences between SOLO and Piaget's stage theory. One is what they categorise. Piaget categorises students by their general level of thinking, but SOLO categorises students' responses. In a Piaget study, the same response or justification could be observed from children at different developmental stages (for example, Piaget & Inhelder, 1975, p. 146 and p. 149). When a student gave responses at different levels Piaget and Inhelder identified the cognitive level of a child based on the level that was most often used in a group of related tasks. Since SOLO shifts the label from the student's intellectual development stage to his or her response to a particular task, such a *cross-level* label is eliminated. Furthermore, this approach has a practical value in school education. The level of response, representing the understanding of the students, provides valuable feedback to teachers and curriculum developers.

The other difference is in the relationship of different levels. In contrast to Piagetian notions, the modes in SOLO theory are cumulative. In other words, the emergence of one mode does not replace its predecessor; the later developing modes exist alongside earlier modes (Collis, 1992). As Rogers (1992) said:

> The modes of cognitive functioning … are not 'stages' which we grow through and out of, and which we abandon as we mature, but 'ways of knowing' which we develop and use as appropriate throughout our lives, and these correspond to the different forms of knowledge we use in our normal daily discourse. (p. 233)

Such a relationship could explain why a child gives different stage responses on different tasks at the same time.

The five basic modes of functioning in SOLO model are sensori-motor, ikonic, concrete symbolic, formal and post formal. As with the Piagetian stages the SOLO modes are increasingly sophisticated. The sensori-motor mode "is well described as tacit knowledge; it consists in 'knowing how' to carry out a skilled act without necessarily being able to describe or explain it" (Collis, 1992, p. 27). In the ikonic mode, oral communication and the skill of using words, images, and so on, are developed and they lead to qualitative judgements. In the concrete symbolic mode, "the concrete world can now be interpreted through symbolic systems" and the elementary systems and physical models "can refer uniquely to events and objects in the environment" (ibid. p. 28). In the formal mode, abstract concepts and propositions are the elements. Theoretical constructs can be manipulated without having an empirical referent. The most sophisticated mode, the post formal implies that individuals "have a complete overview of their discipline, such that they are capable both of advancing knowledge in the discipline and of challenging its fundamental assumptions and structure" (Collis & Biggs, 1991, p.192). This is a level beyond what can be expected of school students.

Agreeing with some other neo-Piagetian cognitive theorists, Biggs and Collis (1991) distinguished features of learning by an hierarchical sequence levels: prestructural (P), unistructural (U), multistructural (M), relational (R) and extended abstract (E).

Studies using Piaget's theory usually focus on children's cognitive stages, not children's misconceptions. Studies using the SOLO framework usually analyse students' responses in detail in order to classify them into different mode and different levels. This certainly does not mean that Piaget's theory is not as useful as non-Piaget's theories, such as SOLO. It means that studies conducted under the two theories have a different orientation. In this study, students' misconceptions and developmental structure are investigated and, consequently, SOLO is selected to be the theoretical framework for this study.

The Application of SOLO in Assessing Students' Responses

SOLO has been used in many topics such as statistics, probability, geometry, arithmetic, problem-solving abilities, and so on. Such research has found its applications in school curricula, instruction, evaluation or describing the development of students' thinking (Collis, Romberg & Jurdak, 1986; Chick, Watson & Collis, 1988; Watson, Campbell & Collis, 1993; Lidster, Pereira-Mendoza, Watson & Collis, 1995; Lam & Foong, 1996; Watson, Collis & Moritz, 1997; Watson & Moritz, 1998).

The SOLO based research most relevant to this study are studies by Lidster, Pereira-Mendoza, Watson & Collis (1995), Watson, Collis & Moritz (1997), and Watson & Moritz (1998). The first paper relates to the part of this study concerned with the interpretation of chance values and the next two papers relate to the part concerned with chance comparison in one-stage experiments of this study.

Originally, Biggs and Collis (1982) described only one U-M-R cycle in each mode. However, in the three studies, Watson and her colleagues identified two U-M-R cycles in the concrete symbolic mode. As this author feels that the original five levels (P-U-M-R-E) suggested in the SOLO taxonomy can satisfactorily describe the different levels of understanding of the concept probability, the two-cycles model was not selected for use in this thesis. However, the finding of the three studies on students' increasingly sophisticated responses to the items is an important reference for the current study.

In Lidster et al. (1995), they described such a developmental process in making a judgement of whether a die is fair. This process can be viewed as follows: own beliefs or experiences → physical properties of the die → a single sample → the die could behave differently in different samples → do more tosses → the information from two or more samples → the trends across a number of samples.

In Watson et al. (1997), they reviewed other relevant research using the marbles problem (Noelting, 1980a, 1980b; Green, 1983a, 1983b; Singer & Resnik, 1992 and Karplus, Pulos & Stage, 1983) and discussed it in relation to the researchers' SOLO work. They indicated that although various levels were suggested and more or less detailed sequences of students' responses were devised in different studies, all the results were consistent.

Watson and her colleagues described such a developmental process of chance comparison in one-stage experiments. This process can be viewed as follows: own beliefs or non-mathematical experiences → anything can happen → qualitative comparison or rudimentary numerical comparison → correct quantitative comparison in a straightforward setting → implicit idea of using ratio in the comparison between two boxes → explicit idea of using ratio in the comparison between two boxes → correct quantitative comparison in a complex setting.

Generally, this author agrees with their classification of the students' responses, but in three cases their coding could be interpreted differently.

First, in Lidster et al. (1995), they coded the following kind of response at the second cycle multistructural level (p. 15): information from two or more samples is put together but lack of recognition that generally a larger sample is a more reliable predictor. From this author's perspective it should be coded at a lower level as the students considered the data in two samples because they were shown information obtained from different samples, not because their thinking contains the elements of abstract thinking such as the more repetitions the more reliable the estimation. Consequently, in this study this kind of response was coded as the multistructural level of the concrete symbolic mode.

Second, for the item, to pick a name out from a hat containing 13 boys' and 16 girls' names, is it more likely that

(b) the name is a boy, or

(g) the name is a girl, or

(=) are both a girl and a boy equally likely?

Please explain your answer.

In Watson et al. (1997) and Watson and Moritz (1998) they labelled the following two responses as unistructural and multistructural response in the first cycle respectively.

U_1: (=), *you could get both.*

M_1: (=), *the chance is the same.*

Although the first response did not involve qualitative comparison in its reason it could be argued that it was implied in the equality answer. It is possible that the real difference that exists between the two responses is not large. In order to refine the ability to code the responses to such items, in this study one more option, "It's impossible to compare the likelihood of the two outcomes" was added.

Third, for the item, two boxes A and B are filled with some red and blue marbles, if you want a blue marble, which box should you choose?

(A) box A (with 6 red and 4 blue)

(B) box B (with 60 red and 40 blue)

(=) it doesn't matter.

Please explain your answer.

In Watson et al. (1997) and Watson & Moritz (1998), they labelled the following response as a relational response in the second cycle.

> R_2: (=), *it doesn't matter for the factor of both boxes are blue over red which is equal to Box A = 4/6 =2/3 & Box B = 40/60 = 4/6 = 2/3.*

However, such a strategy of comparing part-part ratios cannot be generalised to more complicated situations. For example, if marbles of another colour are added to bag A. So the approach cannot be thought of as a correct quantitative comparison generalisation in a complex setting. In this study, only the responses based on comparing part-whole ratios were coded at the extended abstract level of the concrete symbolic mode.

Other research studies that are particularly relevant to this study but not SOLO based are Jones, Langrall, Thornton and Mogill (1997) and Jones, Langrall, Thornton and Mogill (1999). In Jones et al (1997), they suggested a framework for assessing children's probabilistic thinking. Jones et al (1999) extended the framework for assessing both elementary and middle school students' probabilistic thinking. In their research, they incorporate six key constructs (sample space, experimental probability of an event, theoretical probability of an event, probability comparisons, conditional probability, and independence) and four levels (subjective, transitional, informal quantitative, and numerical) for each of the constructs. The framework was validated from the results of their observations and interviews. It appeared to be in general agreement with the other literature. But one situation might need further investigation. In the validation of their primary framework, they found that all the students were able to recognise every instance of certain and impossible events. Then they moved recognises certain and impossible events from transitional level to subjective level (Jones et al. 1997). Such a change is also used in Jones et al 1999. It would seem open to question as to whether all students' responses start from this level. In the literature reviewed here it was reported that other researchers have found students have considerable difficulty with the identification of certain, impossible and possible events. The result that found by Jones et al. (1997) might have been a function of certain components of the study. For example, the marbles of different colours are shown to the students during the interviews rather than just a statement on a written test as was used in other research reported in the literature.

Since there is some evidence that students do have some intuitive strategies in chance measurement and chance comparison and they are also capable of understanding probability, one may accept that students' conceptions and misconceptions of probability can be modified and developed by teaching. Research on practical strategies for probability instruction will be reviewed in the next section.

Research on Practical Strategies

Traditionally the teaching of probability neglected intuitive understanding. The focus was on permutation, combination, arrangement and the classical definition of probability, and this is still the current situation in China. It leads to many students being unable to apply their classroom-based learning to real-world situations. Their knowledge of probability is separate from and sometimes even contradicts the strategies they use in real-life probabilistic situations. The traditional approach had resulted in criticisms of the situation and educators have suggested and implemented reform projects (Newman, Obremski & Scheaffer, 1987; NCTM, 1992; Lovitt & Lowe, 1993;

Interactive Mathematics Program, 1998). A major aim of these projects is to provide opportunities for students to gain experiences in stochastic situations through problem solving or simulation, either manually or by using a computer, before being taught theoretically. These projects attempt to reduce the teaching dependence on algebra, arithmetic, and combinatorics as much as possible and, consequently they use frequentist probability more often than classical probability. Such an approach has also been suggested in China (Study group of school mathematics for the twenty-first century in China, 1995).

Most of the research on students' probabilistic thinking appeared after the 1970s and some common practical strategies for developing students' probabilistic conceptions have been widely suggested.

Hawkins and Kapadia (1984), Garfield and Ahlgren (1988), Konold (1991) and Shaughnessy (1993) recommended some general strategies for probability teaching. Although the recommendations were either generalised from literature, or from their own teaching experiences, some suggestions are common. For example, creating situations to encourage students to examine, modify, or correct their own beliefs of probability or correct others' common misuses of probability by the use of real data, activities and visual simulations. They preferred to base the teaching on students reality to make it meaningful and useful for students rather than abstract and irrelevant. In 1995, in a summary report on what had happened during 1980s and 1990s, Garfield (1995) affirmed these recommendations again but noted the tendency to incorporate technology such as computers, calculators, multimedia, and internet resources in teaching.

Compared to the abundant research on students' misconceptions of probability there is less research on the use of teaching interventions, especially in school level, reported in the literature. For this study, the most directly related teaching experiment is that of Fischbein and Gazit (1984). Their teaching programme (12 lessons) mainly included topics such as the identification of impossible, possible and certain event; the concepts of probability and relative frequency and the relation between them; simple and compound events and their probabilities. Close to 600 students from grade 5,6 and 7 were involved in the study, either in the experimental group (with teaching) or the control group (without teaching). Fischbein and Gazit also believed that students' intuitions cannot be modified by verbal explanations only, so in their teaching programme,

> …. the children were presented with various situations which offered them the opportunity to be active in calculating probabilities, predicting outcomes in uncertain situations, using operations with dice, coins and marbles for watching, recording and summing up different sets of outcomes. (Fischbein & Gazit, 1984, p. 3)

After the teaching intervention they found that about 60%-70% of the grade 6 students and 80%-90% of the grade 7 students were able to understand and use correctly most of the concepts which were taught. For grade 5 students, they could identify impossible, possible and certain events but were unable to understand most of other topics especially chance calculation and the identification of simple and compound events. They found that the teaching also had a positive effect on overcoming some misconceptions that not being discussed in the teaching. But the most important hypothesis they found was that probabilistic thinking and proportional reasoning might be based on two distinct mental schemata. "A progress obtained, as an effect of instruction, in one direction does not imply an improvement on the other" (ibid. p. 23). In an item (estimate the number of fish in a pool) that required proportional reasoning, they found far fewer correct responses than those in other simple computational probability questions. This might hint at the possibility that the two definitions of probability, theoretical and empirical, are also based on two distinct mental schemata.

Shaughnessy (1977, 1992) undertook a teaching experiment at the tertiary level. In his twelve-week, intensive teaching experiment (meeting five days a week for an hour each day) he always started teaching with a specific problem and asked his students to guess the answer. Then the students co-operated in small groups (four or five members), did the experiment, collected and analysed data, and finally reached some theoretical conclusions. He found such a small-group, activity-based, model-building approach can help undergraduate students learn to discover some elementary probability models and formulas for themselves and reduce their reliance upon strategies such as representativeness.

As for practical strategies in overcoming specific misconceptions mentioned earlier in this chapter, Lecoutre and Konold found it was not easy to change students' equiprobable responses and outcome-oriented responses.

The use of equiprobability, as reported previously, when solving simple questions such as lucky draw items decreases with the student's age. However, equiprobable responses arise again in dealing with compound event problems. Lecoutre and her colleagues investigated the resistance of the equiprobability bias used in two-stage compound event problems under a variety of experimental conditions. For example, helping students notice the different ways of getting the same combination (a 5 and a 6); presenting different frequency information by giving the results of 100 and 1000 dice throws; modifying the formulation of the results or the questions asked; helping students utilise and transfer correct probabilistic intuitions to compound event problems. She also reported on their investigations of the influences of students' backgrounds in probability, learning styles and gender. Lecoutre (1992) summarised the results of the studies by saying,

> None of the factors had a strong effect on the equiprobability bias. In almost all experiments, the proportion of equiprobability responses remains at least equal to 50 percent. …even a thorough background in the theory of probability did not lead to a notable increase in the proportion of correct responses. (p. 560)

In addition to the summary, Lecoutre (1992) reported a new experiment. After masking the chance aspect of probability problems (drawing two cards from a bag to construct geometric figure rather than drawing two pieces of candy or two poker chips from a bag), 75% of the students responded correctly. Equiprobability responses massively declined to 23% and the chance model was used by only 5% of the students. However, when the mask was removed, 22% of the subjects used the chance model again. Adding the another 22% of students who consistently gave equiprobable answers to the two items (with and without masking), the proportion of equiprobability responses was still close to 50%.

For the outcome approach, Konold found students' outcome-approach responses "were fairly consistent, both across different problems and a 5-month time interval" (Konold, 1989, p. 87-88). He reported, "The students with low outcome scores in general gave responses that were consistent with a formal interpretation of probability. …Students who had taken a statistics course tended to receive lower outcome scores than those who had not" (ibid. p. 72). A teaching experiment by Konold and his colleague showed that instruction did impact on changing the outcome approach but the effect was small as the correct percentage on the weather problem only increased from 32% to 38% (Konold, 1995).

The theme of the literature on probability teaching is to use real data (obtained from students' own simulations and computer simulations) to help make students aware of their misconceptions, make more informed predictions, and explore and develop formal conceptions of probability. Such an approach is especially appropriate for learning probability, as real patterns are often only revealed after a large number of repetitions. Pooling students' data is an economical way of collecting data. An important point was made by Orton who noted that both co-operative physical activities and independent mental activities are necessary for promoting mathematics learning (Orton, 1994).

Summary

The main difference between probability and other mathematics is that the thinking processes are different. Studies on misconceptions showed that some misconceptions such as the outcome approach (an event with a high probability of happening should happen and if it does not happen, then the assigned probability is wrong) and representativeness (a sample should look like its parent population) are related to causal thinking. They are stable and difficult to change. One possible way to try and overcome the potential problems with such misconceptions is to develop formal views of probability before these misconceptions become firmly established in students' minds. Actually, studies on the development of understanding of probability have shown that it is possible to introduce probability into primary school and secondary school curricula. The teaching should focus on probabilistic ideas and activities rather than on calculations and formulae.

The next chapter will provide an overview of the research methodology and design of the study.

Chapter 3 Research Design and Procedures

This study is divided into two parts. In the first part, referred to as the main study, questionnaires were given to 567 high school students from grades 6, 8 and 12 and a subset of 64 students was interviewed. The purpose of this part of the study was to identify the main misconceptions that students had regarding probability and describe their understanding of probability using a developmental hierarchical structure. The second part, based on an initial analysis of the data from the first part, involved a short-term teaching intervention with two classes at grade 8 in an ordinary school. This chapter outlines the design and procedures used in these two parts. The process of developing and modifying the questionnaires, collecting data, carrying out the teaching intervention and analysing the data is described.

Development of the Instruments

The specific items are described in the next section. Prior to presenting this information, some general comments about the questionnaires and interviews are given.

Use of Multiple-choice Format. Multiple-choice plus explanation was the main format for the items. In order to gather information on students' thinking it is necessary to require students to explain their answers. Given that about two-thirds of the subjects in the main study were junior-high school students without any formal probability training, it was felt that it might be quite difficult for them to answer open-ended questions, so the multiple-choice plus explanation format was used for most items. It is realised that one of the limitations of a multiple-choice format is that students are limited to the pre-selected options. To overcome this, students were allowed to give an answer that did not appear in the options.

Use of multiple questionnaires. The effect of data and context in eliciting students' misconceptions were two dimensions investigated in this study. Consequently, it was necessary to develop parallel test items to reflect these dimensions. The meaning of parallel is that items change only one dimension, data or context. Given the number and parallel nature of the items meant that several sets of questionnaires containing different versions of the "same" item were developed and used in each class.

Assessing students' confidence in their responses. The question of students' confidence in their answers was also investigated in the study. In order to do this, for each question, students were asked to indicate their confidence in their answer.

Design of interviews. The purpose of the interviews was to clarify students' thinking as indicated in their written answers. No specific items were designed for the interviews. The structure was to give the students their written test and ask them to explain their answers to selected items. In addition to the items on his or her questionnaire, as necessary, parallel questions were used in the interview. A subset of approximately 10% of the sample in the main study was interviewed.

Categories of items. The items were developed and assigned to four general categories. They are: Category I - Identification of impossible, possible and certain events; Category II - Interpretation of chance values; Category III - Chance comparison in one-stage experiments; and Category IV

- Chance comparison in two-stage experiments. These four categories are now described.

Category I Items

In this category students were asked to indicate whether an outcome was impossible, possible or certain. The original problems used as a basis for these items came from Fischbein, Nello and Marino (1991). As Chinese students are not familiar with a tombola game, a playing card situation was used instead. It was also felt that there might be some different responses between one-stage and three-stage experiments. For example, consider the confusion between rare and impossible discussed in Chapter 2. Students may believe in a one-stage experiment involving rolling a single die that *rolling a six* is a possible outcome, but believe that in a three-stage experiment where 3 dice are rolled that *rolling three sixes* is an impossible outcome. Consequently, the one-die/one pack of cards tasks that have been widely used by other researchers were extended to include three dice/three packs of cards tasks.

Category II Items

The items in this category focused on the outcome approach and frequentist explanations of probability. The original problem that was used as a basis for these items was Konold's (1989) weather problem.

> What does it mean when a weather forecaster says that tomorrow there is a 70% chance of rain? What does the number, in this case the 70%, tell you? How do they arrive at a specific number?
>
> Suppose the forecaster said that there was a 70% chance of rain tomorrow and, in fact, it didn't rain. What would you conclude about the statement that there was a 70% chance of rain?

> Suppose you wanted to find out how good a particular forecaster's predictions were. You observed what happened on 10 days for which a 70% chance of rain had been reported. On 3 of those 10 days, there was no rain. What would you conclude about the accuracy of this forecaster? If the forecaster had been perfectly accurate, what would have happened? What should have been predicted on the days it didn't rain? With what percentage chance? (p. 96)

The original weather problem that was used with tertiary level students included three sub-problems. There was a concern that if the problem was used in the original format with school students it could be difficult to infer the underlying thinking of the students. Consequently the three sub-problems were separated and adapted, with the first and the last developed into a multiple-choice format. Meanwhile, the items were expanded and adapted to include different contexts and data. The three contexts involved predicting the outcome of a match, predicting the outcome of drawing a marble from a bag and predicting the weather on a given day. The data used were 80%, 50%, and 30%. In all this generated 27 items. Also, when the alternative responses for the multiple-choice items were developed the wording in the choices was varied to study the understanding of uncertainty, frequentist interpretation and the use of the misconceptions such as the outcome approach. Alternatives used with different items included wording such as around 8 out of 10, exactly 8 out of 10, 9 out of 10, 8 out of 10, around 5 out of 10, exactly 50 out of 100 and around 50 out of 100. These foci have received limited attention by other researchers, but are considered an important component in determining students' understanding of probability.

Category III and IV Items

The other two components of the test focused on chance comparison. Category III items were one-stage experiments (such as, drawing one slip from a box or choosing the bag that had the greater probability of having a marble of a given colour). The items were cited in or adapted from Kahneman and Tversky (1972), Piaget and Inhelder (1975) and Green (1982). They were designed primarily to generate information on misconceptions such as chance cannot be measured mathematically, equiprobability and so on. Furthermore, the design of the items allowed for a systematic analysis of the impact of changing data and context.

Category IV items focused on two-stage experiments (spinning the arrowheads of two spinners or drawing one marble from each of two bags). The ability to list sample points in two-stage experiments, the use of misconceptions such as the compound approach (split a two-stage experiment into two one-stage experiments) and the influences of changing data and context were studied. The item to *list all possible results of tossing two coins together* was adapted from Green (1982) and the other items in this part were adapted from Fischbein, Nello and Marino (1991).

In all, using the different contexts, data and categories of items resulted in more than 80 items and they were sent to overseas experts for comment. Based on their feedback some modifications were made to the items to be piloted. Other potential modifications were not made at this stage, since it was felt more appropriate to wait until the results of the pilot provided additional feedback. Finally, while the development of the test items was in English, the tasks were translated into Chinese for administration. The translated version was given to another two experts in Singapore to check the accuracy of the translation from English into Chinese.

The previous section describes the development of items that were used in the study. The following section briefly describes the piloting process that was used to finalise the items.

Pilot Study

Piloting and refining the Items

The items were organised into six different pilot tests. Each set contained approximately 13 items and the order of presentation of the 13 items was changed in different versions of the pilot. The test papers were sent to a colleague in China, together with specific instructions on how to administer the tests. For example, the teacher was asked to show a real die to the students, explain the meaning of a pack of cards without picture cards and not to help the students if they ran into problems. The pilot sample was 144 students in grades 6, 8 and 12 in a small city near Shanghai. The 48 subjects in grade 12 came from an advanced high school and the other 96 subjects in grades 6 and 8 came from two ordinary high schools. The time allowed for the test was approximately 55 minutes.

The results of the pilot study showed the viability of the test in terms of both the time frame and structure. Virtually all students in each grade gave answers and reasons for all the items. However, as would have been expected, based on the results of pilot study together with the original suggestions by the experts some final revisions were made. The changes made were:

1 In the pilot students were not asked to give explanations for their answers to category I items. *In the main study students were asked to give explanations for all items*.

2 Both a qualitative form (no confidence, little confidence, a lot of confidence and full confidence) and quantitative form (no confidence, 25% confidence, 50% confidence, 75% confidence and full confidence) to measure confidence were tested in the pilot. *In the main study only the qualitative form was used*.

3 *In the main study it was decided to focus on four misconceptions, chance cannot be measured mathematically, equiprobability, outcome approach and compound approach.* This meant that the three items developed specifically for positive-recency and negative-recency were eliminated and four additional items related to the compound approach were added. The pilot data indicated that students' judgements in two-stage experiments were worthy of further investigation.

4 The different placements of the arrow in the spinner questions were *not used in main study*. It seems that this was a confounding factor that might have led some students to assume they were required to consider the initial position of the arrowhead.

5 *Some changes in wording and options in the multiple-choice items were made*. This was done to clarify some items as well as to make it easier to identify various strategies that might be used to solve the problems. This clarification meant that some of these items were longer and the language appears more complex when compared to items used in other research.

6 *The options "it's impossible to compare the likelihood of..." and "it's impossible to indicate which one is the most likely among..." were added to the items in category III and IV.* Adding the options was seen as advisable since it allowed students to express simple uncertainty and could be seen as indicating a lower level of understanding of random phenomena.

The Final Items

A total of 83 items was used for the main study (see Appendix A), organised into four categories (see Table 3.1) and nine disjoint questionnaires (see Table 3.2).

Table 3.1 Number of items in each Category used in the main study

Category		Number of test items
I	Identification of impossible, possible and certain events	24
II	Interpretation of chance values	27
III	Chance comparison in one-stage experiments	22
IV	Chance comparison in two-stage experiments	10
	Total	83

Table 3.2 Overview of the nine sets of questionnaire used in the main study

	Category I	Category II	Category III	Category IV
Set A	1(1), 2(2), 4(6)	1(3), 2(4), 3(8)	1(1), (2)	1, 6
Set B	1(2), 2(1), 4(5)	2(3), 3(4), 1(8)	1(3), (4)	8
Set C	1(3), 3(5), 4(1)	3(3), 1(4), 2(8)	2(1), (2)	9
Set D	1(4), 3(6), 4(2)	3(2), 1(6), 2(7)	2(3)	3, 7
Set E	1(5), 4(3), 3(4)	1(1), 2(5), 3(9)	2(4), 5, 9	
Set F	1(6), 4(4), 2(3)	2(1), 3(5), 1(9)	3(3), (4), 6, 10	
Set G	2(5), 3(1)	3(1), 1(5), 2(9)	4(1), (2), 7, 8	
Set H	2(6), 3(2)	1(2), 2(6), 3(7)	3(1), (2)	2, 10
Set I	2(4), 3(3)	2(2), 3(6), 1(7)	4(3), (4)	4, 5

Table 3.3 - 3.6 show the relationships among the items in each category. The arrows joining the boxes illustrate changes in either data or

context. For example, consider Table 3.5. Reading from left to right for the top row, the data is changed. Item III1 Involves 20 Girls and 22 Boys, item III2 involves 5 Girls and 27 Boys and so on. The vertical arrows joining the second and third rows show changes in context from a marble context to a spinner context, while the right arrows show a change in data, either in terms of the number of marbles or the angles between the segments in the spinners.

Table 3.3 Relationships among the 24 items in Category I

I? P? C?	One normal die **I1**	Context →	One pack of playing cards … **I2**	Stage →	Three normal dice **I3**	Context →	Three packs of playing cards … **I4**
(1)	Even Number		Heart or diamond		Even numbers		Hearts or diamonds
(2)	< 7		< Jack		< 7		< Jack
(3)	> 6		> 10		> 6		> 10
(4)	= 2		= diamond 2		= 2		= diamond 2, diamond 5 and spade 8
(5)	= 6		= 10		= 6		= diamond 2
(6)	≠ 6		≠ 10		≠ 6		Different in both number and suit

one-stage (I1, I2) three-stage (I3, I4)

Table 3.4 Relationships among the 27 items in Category II

		Win match **II1**	Context →	Pull out white marble **II2**	Context →	Rain **II3**
(1)	Which one has the closest meaning to the chance of…is…?	80%		80%		80%
(2)		50%		50%		50%
(3)		30%		30%		30%
(4)	The outcome is … Do you think the prediction is accurate or not?	80%		80%		80%
(5)		50%		50%		50%
(6)		30%		30%		30%
(7)	...In which situation, can the prediction be considered very accurate?	80%		80%		80%
(8)		50%		50%		50%
(9)		30%		30%		30%

Table 3.5 Relationships among the 22 items in Category III

	20Gs & 22Bs **III1**	
(1)	The 1st draw	
(2)	2Gs4Bs	The 7th draw
(3)	3Gs3Bs	
(4)	0G6Bs	

Data →

5Gs & 27Bs **III2**	
The 1st draw	
2Gs4Bs	The 7th draw
3Gs3Bs	
0G6Bs	

Data →

400Gs & 440Bs **III3**	
The 1st draw	
15Gs55Bs	The 71st draw
35Gs35Bs	
0G70Bs	

Data →

200Gs & 1000Bs **III4**	
The 1st draw	
23Gs47Bs	The 71st draw
35Gs35Bs	
0G70Bs	

8 red & 16 black
vs.
50 red & 70 black
III5 (want a black)

↓ Context

120° (red) : 240° (blue)
vs.
150° (red) : 210° (blue)
III8 (want stops in blue)

Data →

21 red & 8 black
vs.
210 red & 80 black
III6 (want a black)

↓ Context

270° (red) : 90° (blue) small spinner
vs.
270° (red) : 90° (blue) big spinner
III9 (want stops in blue)

Data →

8 red & 16 black
vs.
500 red & 100 black
III7 (want a black)

↓ Context

120° (red) : 240° (blue)
vs.
300° (red) : 60° (blue)
III10(want stops in blue)

Table 3.6 Relationships among the 10 items in Category IV

2 white & 2 black
vs.
2 white & 2 black
IV1 (list all possible outcomes)

↓ Context

180° (red) : 180° (blue)
vs.
180° (red) : 180° (blue)
IV3 (list all possible outcomes)

2 white & 2 black
vs.
2 white & 2 black
IV2 (compare chances of all possibilities)

↓ Context

180° (red) : 180° (blue)
vs.
180° (red) : 180° (blue)
IV4 (compare chances of all possibilities)

Data

8 red & 16 black
vs.
50 red & 70 black
IV5 (compare chances of all possibilities)

↓ Context

120° (red) : 240° (blue)
vs.
150° (red) : 210° (blue)
IV8 (compare chances of all possibilities)

Data →

21 red & 8 black
vs.
210 red & 80 black
IV6 (compare chances of all possibilities)

↓ Context

270° (red) : 90° (blue) small spinner
vs.
270° (red) : 90° (blue) big spinner
IV9 (compare chances of all possibilities)

Data →

8 red & 16 black
vs.
500 red & 100 black
IV7 (compare chances of all possibilities)

↓ Context

120° (red) : 240° (blue)
vs.
300° (red) : 60° (blue)
IV10 (compare chances of all possibilities)

Main Study

Sample

Clearly it was not viable to select a representative sample from a population spread across China. Both resources and practical considerations limited the actual sample. The sample selected for the main study was 567 students from grades 6, 8 and 12 in Shanghai.

City

As explained in Chapter 1, at the present time most Chinese students do not study probability during their 12 years of primary and secondary schooling. In Shanghai, the situation is the same for the first 11 years of schooling, but there is very limited exposure to probability in grade 12, where the students have a short unit (about 8 hours) involving a formal, theoretical approach to probability. As the plan was to collect data from students who had been and had not been taught probability, Shanghai was selected.

Schools, Grades and Students

Shanghai is a major city in China covering an area of approximately 6300 square kilometres. The population is in excess of thirteen million, with more than seven hundred thousand high school students (usually aged from 11 to 17 years) enrolled in approximately 780 high schools.

Due to the restrictions imposed by the local system it was not possible to randomly select schools and classes. The solution was to select students from both the advanced and ordinary streams that, based on the researcher's experience of the system, would provide a representative picture of students in the system. It was planned to select, where possible, half the classes from each of the streams.

As mentioned in Chapter 1, for the past twenty years Shanghai high schools have been classified into two streams, advanced (containing about the top 10% of students) and ordinary. But from 1997 the classification at the junior level was eliminated. New students usually enrol in a junior high school near their home. The change did not affect the sampling in grades 8 and 12 as the students were enrolled when streaming was still in place and students could be classified as either advanced or ordinary school students. However, since there was no general streaming in grade 6, classes from this grade should be labelled as ordinary. However, as it was the first year after the change, one public advanced school was permitted to select its new grade 6 students from the whole city and carry on its extensive and difficult curriculum in maths. One class in grade 6 was selected from this school and it was labelled as an advanced class.

The subjects of this study were aged from 11 to 17 years. Originally, based on the following considerations, it had been planned to undertake this study in grade 6 (11-12 years), grade 9 (14-15 years) and grade 12 (17-18 years). First, the concept of percentage was initially taught in grade 6. Since percentage was to be used in some items this is the earliest possible grade that can be used for the study. Second, as the initial introduction of probability in Shanghai schools was in grade 12, these are the first group of students who will have had some formal exposure to the topic. Third, grade 9 is in the middle of grades 6 and 12, and data on students between grade 6 and 12 would be the most appropriate in order to observe the development of students' understanding of probability. However, the practical limitation of the situation in Shanghai meant that it was not possible to use grade 9. Grade 9 is the final year of compulsory education in China and students in this grade are extremely busy preparing for the senior high school or other vocational schools entrance examinations. Consequently, grade 8 was selected since it is expected that in the near future the topic of probability will be introduced in junior high school.

Four classes were selected in each grade. Table 3.7 indicates the classes and the number of students per class. The shaded cells indicate the advanced schools. So in grade 6 one class was selected from an advanced school and for grades 8 and 12, half the classes were selected from advanced schools and other half from ordinary schools. The usual class size in Shanghai high schools is over 40 students. One final note on the classes: The grade 12 class 4 students were in an "humanities" programme and students in this programme usually have a weaker mathematics background than the other grade 12 students in this study.

Table 3.7 General information on the 567 students *

	Class 4	Class 3	Class 2	Class 1	Total students' number in each grade
Grade 6	School 3 48 students	School 2 40 students	School 2 36 students	School 1 50 students	174
Grade 8	School 4 54 students	School 4 54 students	School 1 48 students	School 5 53 students	209
Grade 12	School 7 31 students	School 7 56 students	School 6 51 students	School 5 46 students	184

* The shaded cells are classes selected from advanced schools. The students in an advanced school have a better academic background than the students in an ordinary school. Grade 12 classes were the only ones that had received any formal instruction in probability prior to the study.

As was stated earlier in this chapter, it was not possible to randomly select a sample to be representative of all Chinese students. This study can be considered to have a *convenience sample* and, as such, on the surface the results would only pertain to this particular group of students and cannot be generalised. However, based on the following considerations, the researcher feels that the results of this study are still meaningful and valid. The first research question in this study is concerned with Chinese students' misconceptions of probability. Given the nature of the sample it seems reasonable to believe that the main misconceptions of Chinese students should be identified in this study. The developmental structure of students' understanding of probability is investigated in the second research question. Again, it is reasonable to believe that the cognitive framework found in this study should be representative of Chinese students, although the distribution of different levels may be related to the sample. For these reasons, the researcher feels that the use of a convenience sample is acceptable and does not severely limit the conclusions.

Administration of the Questionnaires

The main study was conducted in Shanghai from February 16 to March 3, 1998. In order to maximise consistency, the researcher administered all the questionnaires for the main study, except for the Grade 8 (class 4) and Grade 12 (Class 3) classes since the administration was conducted simultaneously in different classrooms. For these two classes, after the researcher gave a short set of instructions, their own mathematics teachers administrated the test. The instructions were given to each class prior to the administration of the questionnaires. The major points in these instructions were:

- Show them a six-sided die and tell them it has six sides with the numbers 1,2,3,4,5 and 6 on it.
- Show them a pack of cards. Remove all the picture cards to show them a pack of cards without the picture cards.
- Explain that in the multiple-choice items if they did not agree with any of the answers, they could give an alternative.
- Tell them that they are required to give an explanation of their answer for every question.

Forty-five to fifty minutes were allowed for the test, but most of the students finished comfortably within this time limit.

Interviews

After a class completed the questionnaires, the researcher quickly read the written tests and selected a small group of students from that class to be interviewed the next day. The main purpose of the interviews was to check what the students actually meant in their written answers and whether their responses were the results of the misconception implied by the written response.   In many cases the students were also asked parallel problems to clarify their thinking and determine if the thinking was consistent. Other students who gave interesting or unusual responses were also interviewed. A total of 64 out of 567 students was interviewed. Given the practical limitations of the school environment, some interviews were conducted in a corner of a classroom during the lunch break between morning and afternoon classes, so there was considerable background noise. At the beginning of an interview, the interviewee was given his or her own test paper and was told that some answers were unclear that the interviewer wanted him or her to help to clarify the answers. All the interviews were audiotaped.

Data coding

First, an identification code for the form XABCDEFGHI was developed for each student.

X	-	The specific questionnaire, a - i [Main study] j [Teaching study]
B	-	The school, 1 - 7 [Main study] 8 [Teaching study]
CD	-	The grade, 06, 08 or 12
E	-	The class, 1 - 4
FG	-	the student number in the class, 01 - 56
H	-	the sex of the student, Male (1) - Female (0)
I	-	m if interviewed [Main study] All students in teaching study were interviewed so this code was unnecessary.

For example, a5081051 identifies the student took questionnaire A (a), studied in School 5 (5), was in Grade 8 (08), Class1 (1), is the 5th student in the class (05), is male (1), and was not interviewed (no m).

After assigning an identification code to the students, five codes were given to their responses to each item. The first one identified the answer (which choice was selected or what conclusion was made). The second one coded the reasoning (for example, using the outcome approach or not). The third one labelled the cognitive level (for example, level 3). The fourth one noted the agreement between the answer and the reason. The last one recorded the student's confidence level to the solution. Explanations of the coding of the reasoning and cognitive level are given in the next part of this chapter.

Coding of Reasoning

All the expected misconceptions elicited by the test items were listed before the questionnaires were analysed. This initial list was derived from the data obtained in the pilot study and misconceptions that had been identified in the literature. Additional responses were added when the students' responses did not fit the appropriate categories. This resulted in an extensive list of misconceptions. Although it was expected that some of the misconceptions would be related to each other and combined for the purpose of analysis, they were coded separately at this stage. For example, consider item II2(7):

> A mathematician filled a bag with black and white marbles. He didn't really know how many black marbles and white marbles were in the bag. After mixing them thoroughly, he took a look and predicted that "if I pull out a marble from the bag without looking, the chance that it will happen to be white is 80%." Here are five situations. Comparatively speaking, in which situation can the mathematician's prediction be considered very accurate?
> a) pulling out a marble and it happens to be a white marble
> b) pulling out a marble and it happens to be a black marble
> c) suppose that the game is repeated 10 times, white marbles are pulled out 10 times
> d) suppose that the game is repeated 10 times, white marbles are pulled out 9 times
> e) suppose that the game is repeated 10 times, white marbles are pulled out 8 times

The answer of *pulling out a marble and it happens to be a white marble* with the reason that a high probability meant certain to happen was coded as using Konold's outcome approach. The answer of *white marbles are pulled out 10 times in 10 repetitions* with an appropriate explanation (using Konold's outcome approach in each trial) was coded as using the firm outcome approach. While another answer, *white marbles are pulled out 9 times in 10 repetitions* with appropriate justification (applying Konold's outcome approach but with adjustment) was labelled as using the weak outcome approach. The later two answers were both related to the outcome approach misconception but they were kept separate at this stage, as the differentiation might be necessary during the analysis.

Two things should be noted. First, the response to an item was only coded as using a particular misconception when a clear answer and a consistent reason was given. If there was not sufficient evidence for coding, for example, when a response could not be understood clearly or an option selected but there was no clear reason for the choice, even though the option indicated that the student probably using a misconception, the response was not coded as using the misconception. Second, the coding of a response as showing a misconception was only based on the written answers. The information obtained in interviews was not used in counting the number of students using the misconception. Therefore, given the conservative nature of the coding the data on the frequency of a misconception presented in chapter 4 should be less than the real number of students using the misconception.

Coding of Cognitive Level

As indicated in Chapter 2, the SOLO taxonomy was selected as the model. According to the model, student responses to an item can be assigned one of the following levels: prestructural (P), unistructural (U), multistructural (M), relational (R) and extended abstract (E).

Blank, fully irrelevant, illogical, egotistic answers or the inability to become engaged in item answers were coded as P level responses.

If a choice was selected but no reason given, it was coded at the lowest level of the reasons that could lead to the choice. For example, consider item IV6,

> There are 21 red marbles and 8 black marbles in bag A. There are 210 red marbles and 80 black marbles in bag B. Mix the marbles in each bag thoroughly. Put your hands in two bags and pull out a marble from each bag without looking. Which statement below is correct?
> a) it is most likely that both marbles are red
> b) it is most likely that both marbles are black
> c) it is most likely that one marble is red and the other one is black
> d) it is impossible to indicate which one is the most likely among the three outcomes

A grade 8 student, a4084051, in an ordinary school chose (a) but did not give any reason. The lowest level response that could lead to the option is an M level response, such *as 21>8, 210>80, then it is most likely that both marbles are red*. The highest level response could be an E level response that involved using the addition and multiplication rules. So this response was coded as an M level response. It was not treated as missing data because such a situation is different from the situation where the student was not given the task.

Finally, when a student answered a different question due to apparently misreading the item, having difficulty in comprehension or the student apparently was unclear to the researcher intention in the item, and so on, if it was possible to code the answer, it was coded as an answer to the new question. Otherwise, it was treated as missing data. For example, in answering item I2(2) regarding whether the outcome that the card drawn is smaller than a Jack is impossible, possible or certain when a card is drawn from a pack of cards without the picture cards, a grade 12 student, a7124041, misread the word *smaller* as *bigger.* He said it is impossible *since the Jack is* *bigger than the biggest number among 1-10, the probability of drawing out a card that is bigger than a Jack is 0.* His response was coded as an answer to a new question. The response was labelled at the M level as he knew the probability of impossible events is zero.

The data analysis for the three research questions is discussed towards the end of this chapter, after the intervention study is described. However, it should be noted that some preliminary analysis of the data designed to determine the more common misconceptions was undertaken at this stage since this information was used to determine the appropriate teaching intervention.

Teaching Intervention Study

Purpose

One of the preliminary results of the main study was that the interpretation of probability from a frequentist approach was a very weak area for each grade and each stream of students. Many students believed that chance purely depended on luck and cannot be measured mathematically. Therefore, they preferred to interpret probability qualitatively or bet that a certain outcome will happen. They did not think repetition would help. Other students refused to predict the outcome of a trial and chose an option with repetition, but they were not aware of the role of sample size in the fluctuation of an observed frequency. It seems that the frequentist definition of probability is not developed naturally – students need appropriate experiences. In countries where probability has been introduced to young children, the frequentist definition has been widely accepted as the basis for the programme. With the potential curriculum reform involving the inclusion of probability within the Chinese school curriculum, the researcher decided to focus the teaching programme on the frequentist explanation of probability. A short teaching intervention cannot focus all aspects so main objectives of the teaching intervention were to help students understand that

- the specific outcome of each trial cannot be predicted, even after many trials or even when an outcome that has a quite high chance of occurring
- the likelihood of an outcome may be predictable, although whether the specific outcome will occur or not is unknown before the trial
- the larger the number of trials you do, the more reliable your estimation

This teaching approach was primarily designed to help students overcome the misconception that chance cannot be measured mathematically. In addition, the effectiveness of the activity-oriented teaching approach in overcoming students' other misconceptions, such as equiprobability and the outcome approach, was examined.

Subjects

The short-term teaching intervention was carried out with ordinary students in Shanghai in November 1999. Prior to undertaking the intervention, some pilot teaching was undertaken in another school. The pre-test items, post-test items and the workbook were modified based on the pilot experience. Two grade 8 classes, one with 25 students and the other with 26 students, participated in the teaching intervention. The students' ages were 13-14 years except for one student who was 15 years old. Table 3.8 shows the general information on the two classes. The class size was much smaller than usual because this school is a new school, set up only four years ago. With the expansion of Shanghai, more and more residents have moved from the city to new suburbs, so many new schools have been built to meet this relocated population. Like some other new schools, as the student numbers increased the school decided to split the original class in two. This happened about two months prior to the start of the experiment. Another special feature of this school is that a class period is 40 minutes, 5 minutes less than the usual arrangement. This is mainly because it contains classes from grade 1 to grade 12 and one teaching period in primary school is usually 40 minutes.

Table 3.8 General information of the two classes

Class	Girls	Boys	Totals
Class 1 (with computer simulations)	13	13	26
Class 2 (without computer simulations)	16	9	25

As with the main study this was not a random sample. Whatever a researcher does to control the variables, the complexity of the interactions within a classroom makes generalisations from a few classes limited. However, in most studies it is hoped that the situation is representative. Although the class size is smaller than usual, the other aspects of the school situation in this study and background of the students are *typical* and the teaching approach could be replicated in other schools. Consequently, it is felt that for the same reasons discussed in the main study this sample allows the researcher to provide useful information to the educational community.

Pre-test and Interviews

Nine items were used in the pre-test (see Appendix B). Eight of them were selected directly from the 83 items developed for the main study and the other one, relating to the frequentist explanation of probability was new. The criteria for selection were as follow:

- Select items from all the four categories but biased towards category II and III: interpretation of chance values and chance comparison in one-stage experiments
- Select items where students are more likely to use chance cannot be measured mathematically and other specific misconceptions such as the outcome approach.
- Avoid or modify items where there were problems such as misreading or misunderstanding in the main study.

As with the main study, there was a short set of instructions given prior to the test, consisting of the appropriate points mentioned in the main study. In addition, it was explained that there would be six teaching periods on probability. Forty minutes was allowed for the test. Most of the students finished comfortably within this time limit, but some of the slower students did not complete all the items.

Before the teaching intervention each student was interviewed. The purpose and the procedure used in the interviews were the same as that in the main study. In addition, the slower students who did not finish all items on the written test were asked to answer those items orally. The interviews were audiotaped.

Teaching

The teaching intervention involved six activity lessons (40 minutes per lesson) to each class, twice a week. All the lessons used an activity approach (the activities are described in Chapter 6). One class had access to a computer for simulations while the other did not. The class that used the computer had its lessons in a computer lab while the other class remained in its own classroom. At the beginning of each lesson, the students were told what activity would be done that day and what problems they would be solving. Then these problems were discussed briefly with the class. Next, teaching aids such as dice, coins or spinners that were needed for the day's activity were distributed to every student. They were required to do the experiments at least 25 times. Up to this point the two classes, with or without the computer, were taught in the same way. The main difference in teaching was the next part, that is, the way the data for a long series of experiments were generated. The class in the computer lab was exposed to the software *Tangible MATH 3.04 -- probability constructor*, developed by LOGAL Software Inc., while the class without computers received the information from data printed in their workbooks. The students were informed that the data was computer generated.

It should be noted that originally it was planned that the students in the computer lab would do the simulations themselves. Such an approach is preferable to the teacher demonstrating the simulation. However, as a result of the pilot the approach was changed. During the pilot study, the students using computers did not focus on learning probability. The students have **very little prior experience** with computer-based learning, so the students were very excited at having a class in the computer lab. They could not control their desire to play games, draw pictures, and do many other things that had nothing to do with the objective of the lesson. The time, both in terms of the teaching intervention, per se, and time to acclimatise the students to a computer environment were extremely limited. The researcher felt that given these limitations any attempt to have the students use the computer to do the simulations was not viable for effective instruction. Consequently, the less preferable mode of demonstration was selected realising that this might limit the effectiveness of the class with computer instruction. The students looked at the screen on an overhead projector to see the processes of and the results of the simulations undertaken by teacher on the computer. This meant that the difference in the instructional approach between the two classes was less than had been originally expected.

In conclusion the activities, workbook, problems discussed and the teacher were the same for the two classes, with the main difference being that in the computer lab class the data was simulated in front of the students, while in the other class the students were given the computer generated data.

Post-test and Interviews

The instruments used in the pre- and post-test (see Appendix C) were parallel. The test allotted for post-test was 40 minutes, the same to the pre-test, but this time all the students finished in less than the allotted time. Forty-nine students wrote both tests and were interviewed. The two students who missed a test or interview because of illness were eliminated from the analysis.

Coding

The coding paralleled that for the main study, the only difference is that in this case interview data were available on all the students. Because of this the coding was based on the data obtained by combining the questionnaire and interview information.

Data Analysis

The three main research questions are restated and a brief description of the nature of the analysis is given.

Research Question 1

What are the main misconceptions of Chinese students when answering chance interpretation and chance comparison problems?

All the misconceptions observed in the main study were categorised into fourteen groups, with each group focussing on a common underlying conceptualisation. For example, the three misconceptions mentioned earlier, the outcome approach, the firm outcome approach and the weak outcome approach were discussed under one group named the outcome approach. The report for each group usually includes the following parts:

- A description of group
- *Examples*: Each misconception is described and sample student responses are given.
- *Frequency:* Data on how many students in each grade and each stream clearly used the misconception are presented and discussed.
- *Data and Context:* The effect of data and context on eliciting each misconception is discussed.
- *Other:* Additional information from the interviews is presented and discussed.
- *Summary:* A summary of the conclusions for the group.

Most of the analysis was descriptive, based on percentages. For the three main misconceptions, chance cannot be measured mathematically, the outcome approach and the equiprobability, a Rasch analysis was performed to investigate the role of context and data in eliciting the misconceptions. For other misconceptions where the number of students using the misconception was small a Rasch analysis was not undertaken since the small numbers made it inappropriate. A Rasch analysis is a technique for analysing data sets with a substantial amount of missing data. It is particularly valuable when all students in the sample do not answer identical questions, but parallel questions as was the situation in this study (Masters, 1988).

Research Question 2

What is the developmental structure of students' understanding of probability?

More than 95% written responses could be coded into five levels at concrete operational mode of development using the SOLO taxonomy (Biggs & Collis, 1982). The five levels are prestructural, unistructural, multistructural, relational and extended abstract.

The report includes the following parts:

- *Response structure in each category items:* For each category, each level's responses are illustrated and sample student responses are given.
- *Understanding indices:* Data on how many students in each grade and each stream were located at each level are reported and discussed. Then, a Rasch analysis was performed to assign a numerical location to each student.

- *Other:* Additional information from interviews is presented and discussed.

Research Question 3

> Can an activity-based short-term teaching programme improve Grade 8 students' understanding of probability?

As with the other research questions the analysis was initially descriptive in nature, based on the information from the questionnaires and interviews. Students' changes in their answers, cognitive levels and reasoning observed in the two tests are reported. For each item, the analysis includes the following parts:

- A description of the question.
- *Data:* Data on how many students in each class gave correct answers, how many students in each class were located at each level and how many students in each class gave a correct reason and used misconceptions are presented.
- *Discussion:* Students' changes in SOLO levels and reasoning are illustrated and sample student responses are given.

The descriptive data analysis was followed by a Rasch analysis. Each student was assigned two numerical understanding locations – before the teaching and after the teaching. The differences in students' understanding observed in the two tests and the two classes were investigated.

Students' Confidence Level

It has been mentioned earlier that at the end of each item in the questionnaires students were asked to write down their confidence level for their solution. Originally, it was hoped to see whether the students had more confidence when they gave a correct response or a higher-level response. However, based on randomly selected items it was found that the only factor that appeared to impact on the confidence level of responses was the type of item used. Students were generally quite confident in answering category I items (identification of impossible, possible and certain events), less confident in answering category III items (chance comparison in one-stage experiments) and the least confident in answering category II and IV items (interpretation of chance values and chance comparison in two-stage experiments). This means that the students usually felt confident in answering questions associated with familiar situations, such as identifying three kinds of events and less confident in answering in unfamiliar situations. In light of this no further analysis of the confidence data is included in this thesis.

Summary

Data on misconceptions and developmental structure were collected from 567 students in grades 6, 8 and 12. This data were coded and analysed to determine the main misconceptions and the cognitive framework of the students in understanding the concept of probability. Finally, based on a preliminary analysis of the data, a teaching intervention was undertaken with 51 grade 8 students to determine if a short-term activity-based teaching intervention might improve the younger students' understanding of probability and eliminate some of the misconceptions. The next three chapters present the results of the study.

Chapter 4 Main Misconceptions of Probability

Chapter 4 answers the first research question, "What are the main misconceptions of Chinese students when answering chance interpretation and chance comparison problems", and chapters 5 and 6 answer the other two research questions.

Data from 567 students show that there are many specific misconceptions, and they have been grouped together when there was a common underlying relationship, resulting in 14 groups (students could be in more than one group): (1) subjective judgements; (2) example-based interpretations for possible and impossible; (3) possible means certain; (4) chance cannot be measured mathematically; (5) equiprobability; (6) outcome approach; (7) one trial is unrelated to other trials; (8) interpreting chance by data matching or word matching; (9) increasing repetitions is not better for predicting; (10) positive and negative recency; (11) used own methods in chance comparison; (12) taking different order as the same; (13) misuse or extend conclusions inappropriately; and (14) used own methods of chance calculation.

As indicated in chapter 3 the report for each group usually includes the following components:

- A description of the group
- *Examples*: Each misconception is described and sample student responses are given
- *Frequency:* How many students had each misconception
- *Data and Context:* The effect of data and context on eliciting the misconception(s)

- *Other:* Additional information from the interviews
- *Summary:* A summary of the conclusions for this group of misconception(s)

Before discussing the specific results three points should be noted. First, not all the students had the same opportunity to use all the misconceptions. For example, there were only 10 category IV items (chance comparison in two-stage experiments) but 27 category II items (interpretation of chance values). Consequently, the students had more opportunity to use misconceptions related to category II items than category IV items. Second, only those answers with clear evidence of a misconception were coded as using a misconception. This meant that even though the option students selected hinted that they might have used a misconception in solving the problem, their responses were not coded as using the misconception. This is particularly important when the data for the ordinary grade 8 classes were analysed since most of the incomplete answers (a choice was selected but there was no explanation) were obtained from one of these two classes. These two points could lead to an underestimate of the number of students with a misconception. Finally, in this chapter and the chapters that follow there are examples of students' responses to the items. Very minor editorial changes were made to clarify some of the students' responses. However, most are exact translations of the Chinese.

Group 1: Subjective judgements

Although specific questions, such as whether entering the classroom with the right foot will increase students' chances of getting high marks on that day (see Fischbein and Gazit, 1984), that prompt subjective responses were not used in this study, some subjective responses were still observed. Based on their starting point they were divided into two categories. One was *subjective beliefs* and the other *was physical properties based judgement*.

Examples

Subjective beliefs include fully egotistic and other subjective beliefs such as a predicted outcome should happen and a prediction involving a small chance is unreliable. A fully egotistic belief means that instead of considering all the possibilities that exist, students believe the situation can be controlled to lead to a specific outcome or some people are able to predict which outcome will happen. Physical properties based judgement means that whether a specific outcome is going to happen is mainly due to physical properties such as a spinner's size, weight or the starting position of an arrowhead, and so on.

<u>*Subjective Beliefs*</u>

Here are some examples of fully egotistic beliefs. Consider item I1(4): Is it impossible, possible or certain that the number rolled is 2 when you roll a die? A grade 8 student, d1082170, chose *impossible* and wrote that *according to my experience, the more I want a number the less chance it is rolled*.

In a similar item, I1(5), students were asked about rolling a 6. A grade 12 student, e7123260, argued she was sure to roll a 6, *as that was the number I wanted in my mind.* She believed that wishing for something could make it happen.

A grade 6 student in an ordinary school, e2063161m, consistently used the misconception in answering category II items. For item II3(9) where he was asked under what conditions a forecaster's prediction that tomorrow there is a 30% chance of rain could be considered as very accurate, he replied,

Answer: *(a) it really rains tomorrow*
Reason: W*hat a weather forecaster says should be always accurate*.

In the interview, he was asked to answer the same item with the 30% changed to 80%. He said he would still chose that option and gave the same reason. When asked to answer two other items, II1(1) and II2(5), he applied the same strategy and said that an 80% chance of winning a game means that the team will certainly win *as what a coach said is always right* and he believed the mathematician's prediction was accurate, *as he (the predictor) is a mathematician*.

In this study another two beliefs were observed. The first one is that a predicted outcome should happen independent of whether it had a large or a small chance of happening. For example, item II2(3) asked students to select from the following choices the one that had the closest meaning of a 30% chance of pulling out a white marble.

a) the marble pulled out will certainly be white
b) the marble pulled out will certainly be black
c) suppose that the game is repeated 10 times, white marbles are pulled out around 3 times
d) suppose that the game is repeated 10 times, white marbles are pulled out exactly 3 times

A grade 12 student, b5121071m gave the following answer and reason:

Answer: *(a)*
Reason: *As what he said is white marble not black marble.*

In his interview when the percentage 30% was changed to 80% and 10% he still chose (a).

The second belief is that if a prediction involves a small chance, then it is unreliable. A grade 12 student, e7123260, expressed the belief in answering item II3(9) that has been mentioned previously. She said that since *30% is even lower than 50%, it indicates the weather forecaster's report is unreliable.*

These students form a subset of the students who are labelled as using subjective beliefs.

Physical Properties Based Judgements

In these responses students usually talked about weight, angle, strength, time and speed that they believed were related to the outcome of a trial. For item III9, a grade 8 student, e5081260, was asked to predict which spinner, the smaller one or the bigger one, had a greater chance of stopping in the blue part. She concluded that they were the same since *the two angles between the arrowheads and the blue parts on the two spinners are almost the same*. For the same item, a grade 12 student, e7123231, preferred to select the bigger spinner since *it has a bigger surface and if the same strength is used, its arrowhead might be still in the blue part, but the arrowhead of the smaller spinner might be across the blue part*.

Frequency

Subjective judgements were not observed very often in this study. As the two subjective beliefs that a predicted outcome should happen and a prediction involving a small chance is unreliable are not fully egotistic in nature they are counted separately under the general heading of subjective beliefs. Table 4.1 contains the number of students who made subjective judgements.

Table 4.1 Number of students with subjective judgements

Total number of students in each stream of each grade	Grade 6*		Grade 8*		Grade 12*		Total
	Adv	Ord	Adv	Ord	Adv	Ord	
	50	124	101	108	97	87	567
Subjective beliefs (*Fully egotistic beliefs*)	5 (10) *0*	6 (5) *1*	7 (7) *2*	4 (4) *2*	3 (3) *1*	5 (6) *3*	30 *9*
Physical properties based judgement	4 (8)	17 (14)	5 (5)	4 (4)	3 (3)	2 (2)	35

* Figures in italics represent the number of students with fully egotistic beliefs. Figures in brackets represent the percentage of students with the specific misconception.

Fully egotistic responses were observed in nine students' answers, less often than the responses of the two subjective beliefs, and far less often than physical properties based judgements. Nearly half of the physical properties based responses were made by ordinary grade 6 students. It seems that the two subjective beliefs and the physical properties based judgements decrease with age. In grade 6, more subjective beliefs were observed in advanced school students' responses and more physical properties based judgements were observed in ordinary school students' responses. The number of responses in grades 8 and 12 were small and there is no pattern that clearly differentiates advanced schools from ordinary schools.

Data and Context

Regarding the effects of data and context there is no pattern in subjective beliefs based judgements. However, the effect of context was obvious in eliciting physical properties based judgements. Thirty-four of the 35 responses were observed on spinner items. Among the spinner items, the two involving smaller and bigger spinners elicited the most physical properties based judgements.

Other

Usually a physical properties based judgement led to a definite decision regarding the most likely event, but this was not always the case. Three of the 35 students who considered the physical properties of spinners indicated that it was impossible to determine which outcome was most likely. All three responses were to item IV4. One of the students, a grade 12 student in an advanced school, i5121431m, was interviewed. In the interview, he explained his answer this way:

> Student: *From the picture, the two spinners and the positions of two arrowheads are the same. A person's strength is also no big change. So each of options a, b and c could happen. It seems no strategy can be used to find out which one has the biggest chance.*
>
> Interviewer: *Do you mean the three outcomes are equally likely?*
>
> Student: *No, both arrowheads stopping in red part or stopping in blue part are more likely.*

He could not indicate the most likely event because he could not tell which one was more likely between the two events (both in red or both in blue). However, his thinking was still based on the starting position of the arrowheads.

Summary

Although subjective judgements based on subjective beliefs or physical properties were not observed very often in this study the following conclusions can be drawn. First, Chinese students do use subjective based judgements in answering probability questions. The second, while the small number of responses make any conclusions tentative, it appears that physical properties based judgements are dependent on context and seem to decrease with age.

It is important to note that subjective misconceptions may be more prevalent than the data implies. The design of the items did not focus on eliciting this response. Realistically it was only the spinner problem that was likely to elicit physical properties based thinking.

Group 2: Example-based interpretations for possible and impossible

This misconception appears in students' responses to category I items where they were asked to choose whether a particular event associated with a card(s) or die (dice) was impossible, possible or certain. The underlying misconception is a result of students misusing examples as a basis for their decision making. For example, students thought that finding a specific example where an event did not occur meant that the event was impossible. Other students thought that finding an example or examples of an event occurring meant that is was possible and did not consider that, depending on other outcomes, the event might be certain.

Examples

Example-based interpretations for **possible** use one, several or even all outcomes to support the conclusion that it is possible for the target event to happen. Student h4083480, a grade 8 student in an ordinary school, gave the following answer to item I3(2) which involved rolling three dice with the result that all the numbers were less than 7. She selected *possible* as her answer giving the reason that *all the three numbers rolled could be 1.* She found an example but did not realise or consider that while a single example can show that the event is possible, such a strategy does not enable her to determine whether an event is certain.

Another example was given by a grade 6 student, d3064200. When answering item I4(2) which asked about the event that three cards (each drawn from a different pack of cards without picture cards) are all smaller than a Jack, she wrote:

> Answer: *Possible*
> Reason: *In a pack of cards without the picture cards, all the cards left are smaller than a Jack, as the Jack, Queen and King have been taken away. So no matter what card you pick out, it is smaller than a Jack.*

She considered all possible outcomes of the experiment and found all of them supported the target event happening but still concluded that the event is possible rather than certain. This rationale seems to parallel the explanation given by Fischbein et al (1991) when they said that if a certain event is associated with multiple possible outcomes, "the notion of *possible* comes naturally into mind" (p. 527).

Example-based interpretations for **impossible** always use one or several examples to support the idea that it is possible that the target event will not to happen and then conclude that the event is impossible. A grade 6 student, d3064181m, was given item I1(4) asking him whether it was possible, impossible or certain to obtain a 2 when rolling a die. He replied

> Answer: *Impossible*
> Reason: *Because if you are lucky, you may roll out 6 or 5 ….*

He implied that since it was possible to get an outcome other than a 2, rolling a 2 was impossible.

Student d3064200, whose response has been mentioned previously applied the same strategy in interpreting impossible. When answering the same item I1(4) she wrote

Answer: *Impossible.*
Reason: *This is because when the six-sided die falls down on an object it turns to a number, such as 1, 2, 3 … each of them is likely. It's not certain that the number rolled is 2.*

Although both supportive and non-supportive examples were included in her reason, she emphasised that it was possible that the target event might not occur. Example-based reasoning led the two students to conclude that a possible event was impossible.

The following examples are different, either in terms of their answer or reason, but they were labelled as using example-based reasoning. They are grouped under four headings.

Instead of giving specific examples they use the word luck. A few students who chose possible or impossible mentioned the word luck in their explanation but did not use examples. For example, a grade 6 student, c2062140, answered I1(3) about rolling a die once and getting a number bigger than 6. She selected *possible* and her reason was that *I think it depends on one's luck.* The information from these students' written responses was limited. However, as some students like d3064181m who gave specific examples also talked about luck, these ***few*** responses without specific examples were still labelled as using the misconception.

Both possible and impossible were ticked. Another group of students used example-based reasoning but concluded that the target event was both possible and impossible. For example, in answering item I1(1) which asked about getting an even number when you roll a die, a grade 8 student, a4084011m, chose both *impossible* and *possible* and gave the following reason, *because sometimes possible sometimes impossible, but not certain.*

None of the given options were ticked. Instead of ticking both, most of the other students did not tick any options and gave their own options, such as *uncertain* or *sometimes it occurs, sometimes it does not occur*. For example, for item I1(1) a grade 6 student, a3064051, from an ordinary school answered:

> *The correct answer should be uncertain. This is because, if there are four players. Each person rolls the ordinary die once. Four possibilities are the following: The first person rolls a 4, the second person rolls a 3, the third one rolls a 1 and the last player rolls a 6. You see everybody gets a different number. So it is uncertain the number rolled is an even number when a cubic die is rolled.*

For item I3(4), which asked the students to consider the outcome that all three numbers rolled are 2, his classmate, e3064251, answered in this way:

> *The correct answer should be most of the cases impossible, few of the cases possible. For three dice, the likelihood is rare after all. However, it doesn't mean without possibility. So I choose most of the cases impossible and few of the cases possible.*

It appears that the students interpret each outcome of the experiment as separate events. Thus, if they find some outcomes that support the conclusion that the target event may happen and others that support the conclusion that it may not happen, they consider the target event as both possible and impossible.

Correct answer but with example-based reasoning. In the previous examples this reasoning led to incorrect answers. However, a very few students used this reasoning but ticked the correct choice. For example, a grade 6 student, a2062010, said rolling an even number was *possible*, but her reason was because *both impossible and possible are possible.*

Frequency

Table 4.2 Number of students with the example-based misconception in 24 Category I items

Total number of students in each stream for each grade		Grade 6*		Grade 8*		Grade 12*	
		Adv	Ord	Adv	Ord	Adv	Ord
		50	124	101	108	97	87
Number of students who used the misconception	0 out of 2 or 0 out of 3	43 (86)	94 (76)	97 (96)	96 (89)	96 (99)	86 (99)
	1 out of 3	5 (10)	15 (12)	3 (3)	7 (6)	1 (1)	1 (1)
	Half or More than half	2 (4)	15 (12)	1 (1)	5 (5)	0 (0)	0 (0)

* Figures in brackets represent the percentage of students using the misconception for the corresponding cohort.

Table 4.2 contains the number and percentage of students who used the misconception in answering items in category I. Six questionnaires each contained three items in the category and the other three questionnaires each contained two items in the category. So “0 out of 2 or 0 out of 3” in the table means no example-based reasoning was observed in the two or three items answered.

It seems this misconception mainly existed in younger ordinary school students and it decreases with age, virtually disappearing in grade 12. Since the number of students using the misconception is small, the role of data and context cannot be examined.

Other

Some students were interviewed as it appears from their written answers that they had misread the items. When they were required to answer the items again example-based reasoning was observed. Consider a grade 6 student, d2062160m. She was given item I4(2) regarding the event that all three cards drawn were smaller than a Jack. Because she had misread “J”(ack) as the number "1" she drew the wrong conclusion, namely impossible. In the interview, she was asked to answer the item again.

Student:	*Possible, as Jack is fairly big*.
Interviewer:	*Could you give me any examples that the cards are bigger than a Jack?*
Student:	*The Queens and Kings are bigger than a Jack.*
Interviewer:	*But all of them have been taken away.*
Student:	*Then all the three cards are smaller than a Jack must be possible*.

This is another example of choosing possible rather than certain.

Why do students consider all possibilities for a certain event but still conclude that the event is possible rather than certain? A reason was suggested by Fischbein et al (1991) and supported by Chan (1998) and this was discussed in Chapter 2. However, the answer for another question is unclear. Why do students give an example that supports the conclusion that an event is possible but still conclude that it is impossible? The thinking might seem surprising. Based on the data of this study, the researcher suggests two possible justifications.

The first is that the students’ concept of impossible allows for some possible but unlikely events. Consider a grade 8 student, a4083010m, in an ordinary school who answered all three items with impossible (all wrong). When interviewed she emphasised that both occurring and not occurring was possible, but she still felt they were unlikely to happen. She admitted unlikely is different from impossible, however, she chose impossible. The reason was even more clearly stated by another grade 8 student during the teaching experiment (the experiment will be discussed in chapter 6). She argued that any outcome with a chance less than 80% was taken as unlikely to happen at the first trial and should be labelled as impossible.

The second is that they had difficulty in understanding the items. They might have interpreted the question of possible, impossible or certain to mean whether the target event will occur in a single trial. So having positive and negative examples means that it is impossible to determine the result or it is impossible to say the target event will certainly occur in the trial. In the interview, the grade 6 student, d3064181m, who has been mentioned earlier insisted that the outcome of rolling a 2 is impossible since it is impossible that you will always roll a 2; other numbers can be rolled as well.

Summary

Example-based reasoning was a main misconception observed when students were answering items relating to whether an event is impossible, possible or certain. It was mainly used by younger students and ordinary school students and had virtually disappeared in grade 12. Students using this approach gave particular example(s) as their evidence for judging whether an event was possible or impossible.

Group 3: Possible means certain

Another misconception observed is that students believe if it is possible that an event can happen then it will certainly happen, even if there is only a small chance.

Examples

All the responses in this group emphasised the target event could happen then concluded the event is certain.

A grade 6 student, a2062020, gave the following answer and reason for item I1(1) where she was asked whether it is impossible, possible or certain that when a die is rolled once the number is an even number.

Answer: *Certain*
Reason: *Because it is possible the number rolled is an even number.*

For the same item a grade 12 student, a6122061 also chose certain and included a calculation of the probability in her reason:

$$P(A) = C_3^1 / C_6^1 = 3/6 = 1/2 .$$

Even for a rare outcome, such as rolling three 6's, another grade 12 student, c5121131m, in an advanced school chose certain and said that *the situation will certainly happen because the outcome exists.*

Frequency

This misconception is not common but it is unusual as it is mainly applied by good or older students. Five out of the 24 students supported their decision by working out the probabilities of the target events. This result means that although they know how to calculate the probability of an event, they cannot distinguish possible from certain events.

Table 4.3 Number of students indicating that possible means certain

Total number of students in each stream of each grade	Grade 6		Grade 8		Grade 12	
	Adv	Ord	Adv	Ord	Adv	Ord
	50	124	101	108	97	87
Possible means certain	1	5	5	1	7	5

Data and Context

There are 16 items used in this study where the answer is possible and could elicit the misconception. This misconception was observed in 13 of these items. About half of the 24 students used the misconception whatever the data or context was, so even with the limited evidence of the misconception it seems reasonable to conclude that there is no effect due to data or context.

Other

According to the students' answer sheets, one-third of the 24 students changed their option from possible to certain. It seemed this misconception might not be *fixed* in some students' thinking. This is supported by the results of the interviews with two grade 6 students and one grade 8 student, f1061321m, a2063011m and d5081181m. All of them changed to the correct answer and gave a valid reason after being reminded that the target outcome was for one trial.

However, changing their thinking may not be easy for all students. Consider the case of a grade 12 student, c5121131m, whose response has been mentioned previously in this section. He used the misconception consistently in answering items I3(5) and I4(1). In the interview, after being reminded that the target outcome in item I3(5) was for one trial, he said the target event might happen for one trial. When he was asked whether he wanted to change his answer, he said that *perhaps the option possible is better*. When the interview moved on to item I4(1) he still chose certain and said the item was similar to the last one. The researcher asked him that since the two items were similar why had he changed his answer for the first item but not for the second one. He remembered and said after hesitating that *if you only draw once, although such an outcome is somewhat certain, perhaps possible is better*. Clearly he was not fully confident of his new answer.

This unusual misconception might also seem surprising. You roll a normal die and ask students whether you can get a 2. The students indicate that it is possible to roll a 2, and might even tell you that the chance is 1/6, but still conclude that the event is certain. There are two possible explanations for this. One explanation is they think that *although I am not so lucky as to get the target outcome in my first trial, it is certain there is someone who will get it in the first trial*. This reason was given by a grade 8 student in the teaching experiment.

The other explanation is they think the word certain is better than the word possible in describing an event that will definitely happen sooner or later. The fact that one-third of the 24 students changed their first, correct answer of *possible* to their final, wrong answer *certain* supports this explanation. Their overemphasis on the possibility that the target event will happen might lead them to ignore the possibility that it will not happen. So they think may happen/may not happen (possible) is not as good as it is sure to happen (certain).

Summary

Students with this misconception indicated that possible events are certain events since they will definitely happen. In their reasoning, they emphasised the possibility of the target event. This misconception was mainly observed in good or older students' responses.

Group 4: Chance cannot be measured mathematically

For many students or even adults with little understanding of probability, chance means the same as luck. It leads to a misconception that chance cannot be measured or predicted mathematically. This misconception usually led to the conclusion that it is impossible to compare the likelihood of all the possible outcomes.

Examples

Since the students believe that chance is wholly dependent on luck, any attempt to measure it or predict it is unnecessary or meaningless. Here are some examples from the students' responses.

Item IV2 posed a situation that there are two bags each containing 2 white and 2 black marbles. The students were asked to select a correct statement from the following options:

a) it is most likely that both marbles are white
b) it is most likely that both marbles are black
c) it is most likely that one marble is white and the other one is black
d) it is impossible to indicate which one is the most likely among the three outcomes

A grade 8 student, h5081471, replied in this way:

Answer: *(d)*
Reason: *All the first three have a bit of a possibility. Actually, the game basically is taking a chance. The so-called likelihood is non-existent. Which two you draw out is just the two you get.*

Item III4(1) involved the situation of picking out a slip from a box that contains 200 girls' names and 1000 boys' names. The students were asked to select a correct statement from the following options:

a) it is more likely to pick out a boy's than a girl's name
b) it is less likely to pick out a boy's than a girl's name
c) it is equally likely to pick out a boy's as a girl's name
d) it is impossible to compare the likelihood of the two outcomes

A grade 6 student, g2063290 gave the following answer and reason:

Answer: *(d)*
Reason: *How can you know whether he will pick out a girl's or a boy's name. I think it's completely unnecessary to compare, because how can you explain if it happens to be a girl's name, but we superficially think boys have more chance as there are more boys?*

For the same item, g1082330, a grade 8 student in an advanced school gave the following answer and reason:

Answer: *(d)*
Reason: *Since the condition mix thoroughly has been given, the question of likelihood does not exist any more.*

Another grade 8 student, h4084440, in an ordinary school used similar thinking repeatedly on all chance comparison items in her questionnaire. She said that *it is impossible to make a comparison since luck cannot be predicted and is uncertain.*

Some students believe that for the same outcome different trials can have a different chance. This might be part of the reason that led to the conclusion that chance cannot be measured, as the following student's reply illustrates. For the same item III4(1), a grade 12 student, g6122391 wrote:

Answer: *(d)*
Reason: *This is because the principal only picks out one slip. For all the [observed] outcomes [the slips picked out], their chance is the same, equals to 100%. Ratio is only meaningful for many trials.*

The student had been taught probability and in the item the difference between the number of girls and boys was very large, 800 students (200 vs. 1000). He believed that probability was meaningful for many trials, but for one trial it was either 100% (for the observed outcome) or 0 (for all the other outcomes that did not occur in the trial). Since he was asked the chance for a specific draw, he thought the chance could not be measured by ratio.

Other group of students took to predict which outcome is more likely to happen to mean to predict which outcome will happen. For example, for item III2(1), a lucky draw item similar to item III4(1) mentioned previously, a grade 6 student, c3064151, gave the following answer and reason.

Answer: *(d)*
Reason: *As the slip can be a boy's name or can be a girl's name. In the case a girl's name is picked out, it means that a girl's name is more likely; in the case a boy's name is picked out, it means that a boy's name is more likely. So it is impossible to compare.*

In item III3(1), the number of girls and boys is close, 400 vs. 440. A grade 8 student, h5081461 gave the following answer and reason:

Answer: *(d)*
Reason: *Although boys are 40 more, if you put them together, the total number is more than 800. 40 names is only 1/20, it just like a drop in the ocean, almost no influence. So it's impossible to compare the likelihood.*

Then he was given item III3(2) where he was asked what would happen after 15 girls' and 55 boys' names were picked out with no replacement. He answered:

Answer: *(d)*
Reason: *After these names were picked out, the number of boys and the number of girls is equal. What name can or cannot be picked out wholly depends on luck and chance. Further, there are so many names in it, it is definitely impossible to predict.*

Another grade 6 student, a3064040, answered two parallel items III1(1) and III1(2). She gave a correct answer and reason to the first item but answered the second item, when the numbers were equal as follows:

Answer: *(d)*
Reason: *Now the number is same so it's impossible to make a comparison.*

Faced with a chance comparison task all these students thought they were required to predict the outcome. Since the outcome is uncertain and due to luck they seemed to conclude that it is impossible to predict which outcome is the most likely one. Students often emphasised that each outcome is possible and this makes chance comparison impossible or very difficult. Such thinking was elicited slightly more frequently when the numbers for each part were close or the same.

Frequency

This misconception was one of the most common misconceptions observed in this study. Items in category I only related to identifying impossible, possible and certain events and could be answered correctly without quantifying chance. Items in category II focused on the interpretation of chance values. Chance values were given in these tasks and this should reduce greatly the opportunities to use the misconception. Actually there were only 4 students who used this misconception only in category II but not in category III or IV. Categories III and IV were designed for chance comparison and most responses using the misconception were observed in these two categories and, consequently, it is only data from these two categories that is given in Table 4.4. Among the nine questionnaires in this study, four sets have 4, four sets have 3 and one set has 2 chance comparison items from the two categories.

The data shows that this is a common misconception. 46% of grade 6 students, 35% grade 8 students and 19% of grade 12 students used this misconception at least once. 28% of the grade 6 students, 19% of grade 8 students and 8% of grade 12 students use the misconception in at least half of the chance comparison items. Within each stream it seems that the use of the misconception decreases with age, except for the slight increase from grade 6 to 8 for advanced school students. Between streams, ordinary school students use the misconception at least as often as advanced school students in grades 6 and 12 (more often in 6 cells and equal in one cell of Table 4.4) but the trend is in the other direction in grade 8. This is a misconception that is common in grade 6 and 8 but still exists for a significant minority in grade 12.

Table 4.4 Number of students with the misconception that chance cannot be measured mathematically on the 30 chance comparison items

Total number of students in each stream for each grade		Grade 6*		Grade 8*		Grade 12*	
		Adv	Ord	Adv	Ord	Adv	Ord
		50	124	101	108	97	87
Number of students who used the misconception	0 out of 2; 0 out of 3; 0 out of 4	34(68)	60(49)	60(59)	77(70)	83(86)	66(77)
	1 out of 3 or 1 out of 4	7(14)	25(20)	18(18)	15(14)	8(8)	13(15)
	1 out of 2 or 2 out of 4	2(4)	14(11)	9(9)	6(6)	4(4)	3(3)
	2 out of 3 or 3 out of 4	4(8)	17(14)	5(5)	4(4)	0(0)	2(2)
	2 out of 2; 3 out of 3; 4 out of 4	3(6)	8(6)	9(9)	6(6)	2(2)	3(3)

* Figures in brackets represent the percentage of students using the misconception for the corresponding cohort.

Data and Context

Items in categories III and IV were designed to vary along two dimensions, data and context. The three contexts were picking out a name, picking out a marble and spinning an arrowhead of a spinner. The three types of data were equal, close and far apart.

Table 4.5 Use of chance cannot be measured mathematically on the 30 chance comparison items

Item	Grade 6*		Grade 8*		Grade 12*	
	Adv	Ord	Adv	Ord	Adv	Ord
III1(1)	4/9	6/14	2/11	2/13	0/12	0/6
III1(2)	3/9	3/14	1/11	2/13	0/12	0/6
III1(3)	1/7	4/13	1/12	3/12	0/11	0/5
III1(4)	0/7	1/13	1/12	2/12	0/11	0/5
III2(1)	0/5	2/15	3/9	3/12	0/11	1/12
III2(2)	0/5	1/15	1/9	0/12	0/11	1/12
III2(3)	0/5	0/13	2/11	0/12	1/9	1/12
III2(4)	2/5	4/14	2/12	1/12	0/10	0/10
III3(1)	0/5	7/14	5/12	2/11	0/11	3/10
III3(2)	0/5	4/14	4/12	1/11	0/11	3/10
III3(3)	0/4	5/15	0/11	2/12	2/11	0/10
III3(4)	0/4	5/15	0/11	1/12	2/11	0/10
III4(1)	0/5	6/14	1/11	1/12	1/12	0/10
III4(2)	0/5	4/14	0/11	1/12	1/12	0/10
III4(3)	0/5	4/12	3/12	1/12	1/10	1/12
III4(4)	0/5	4/12	3/12	2/12	1/10	1/12
III5	2/5	5/14	0/12	1/12	0/10	1/10
III6	0/4	3/15	0/11	2/12	1/11	0/10
III7	0/5	4/14	1/11	0/12	0/12	0/10
III8	0/5	4/14	0/11	0/12	1/12	0/10
III9	1/5	3/14	2/12	0/12	0/11	0/10
III10	0/4	0/15	1/11	0/12	2/11	0/10
IV2	2/5	12/14	5/12	6/11	1/11	5/10
IV4	2/5	11/12	9/12	4/12	1/10	6/12
IV5	2/5	9/12	7/12	3/12	0/10	2/10
IV6	3/9	2/14	5/11	2/13	0/12	0/6
IV7	1/5	1/13	2/11	3/12	2/9	1/12
IV8	3/7	4/13	4/12	4/12	3/11	1/5
IV9	1/5	2/15	6/9	3/12	1/11	4/12
IV10	0/5	5/14	5/12	2/11	0/11	3/10

* a/b (# using the misconception in grade)/(Total # writing item in grade)

Table 4.5 which reports the frequency of using the misconception in the 30 chance comparison items was used to "eyeball" the data to see if it appeared that the data or context played a role in students using the misconception. For example, is a student more likely to use this misconception when faced with an item where the number of boys and girls was close (item III1(1)) as compared to a parallel item where the number was far apart (item III4(1))? Does it matter whether the item involves marbles (item III7) or spinners (item III10)?

An analysis of Table 4.5 gives some indication of the variability associated with the results on individual items. It seems that the 1st draw items (III1(1), III2(1), III3(1) and III4(1)) are more likely than others (the 7th or the 71st draw items) to elicit the misconception. However, since each student only answered two, three or four of the 30 items it is difficult to "eyeball" the data to determine if there are any real underlying patterns or whether the results are an artefact of different students answering different items. However, if each student's responses on the two, three or four items were coded as either using or not using the misconception and the other 28, 27 or 26 responses coded as missing data, Rasch's Dichotomous model can be used to determine the 30 items locations in eliciting the misconception. For this reason, a Rasch analysis of the data was undertaken. This information is contained in Table 4.6.

Table 4.6 Location Index for 30 chance comparison items in eliciting chance cannot be measured mathematically responses

Item	Location		Item	Location		Item	Location
III1(4)	*		III7	-0.430		III3(4)	0.409
III2(3)	-1.906		III8	-0.430		IV5	0.552
III2(2)	-1.621		IV10	-0.411		III1(1)	0.680
III4(3)	-1.601		III4(2)	-0.214		III3(3)	0.723
III1(3)	-1.426		III3(1)	-0.031		IV2	1.132
III4(4)	-1.237		III6	0.005		IV9	1.461
III3(2)	-0.975		III2(1)	0.020		III4(1)	1.488
III10	-0.805		IV6	0.126		IV8	1.506
III1(2)	-0.749		III2(4)	0.272		IV4	1.626
III9	-0.487		III5	0.338		IV7	1.986

* Item III1(4) was deleted as an extreme item by the computer when running the analysis.

Before discussing this table it is important to explain the meaning of item location. It is not a measure of whether or not the item was answered correctly. It is a measure of the difficulty associated with eliciting this misconception as the rationale for the solution. The higher the location of an item the more likely it was to elicit the misconception.

Six out of eight of the category IV items were in the highest location, indicating that category IV items were more likely than category III items to elicit the misconception. This might be due to the wording of the items in this category. Although answers that clearly stated I don't know were not coded as this misconception, some students might still have considered the option that it is impossible to indicate which one is the most likely among the three outcomes as synonymous with I don’t know.

An analysis of the results in Table 4.6 implies that context did not play a role in eliciting the misconception. Items III5, III6 and III7 were designed to be parallel to items III8, III9 and III10, but no pattern was observed. A similar conclusion applies to comparing the results of items IV2, IV5, IV6 and IV7 with items IV4, IV8, IV9 and IV10.

There is also no clear evidence relating to the role of data. Changing the total from a small number (about 40 slips, items III1) to a large number (about 1200 slips, items III4) or changing from close numbers (400 girls vs. 440 boys, items III3) to very distinct numbers (200 girls vs.1000 boys, items III4) resulted in no major shift in the number of students using the misconception. The only pattern is that items asking about the likelihood of the first slip tended to elicit the misconception slightly more often.

Other

As mentioned earlier, about 20% of the grade 12 students used the misconception at least once in two, three or four items. They had had a short teaching unit of probability several months before the questionnaires were given. So some students using this misconception were able to use their theoretical knowledge to calculate the probability of an outcome occurring. However, they doubted it was meaningful or could be taken as a guide when making a decision in the real world. Their real-life experience prevailed. For example, a grade 12 student, d7123121m, was interviewed. He said if the questionnaire has used the term probability instead of likelihood he probably would have answered in a very different way. This is because

> *Likelihood does not equal probability*. *Likelihood is associated with the real life problem but probability belongs to mathematics. My mathematics teacher told me that the probability of getting a head is 1/2 when a fair coin is flipped. However, if she is asked to make 1000 flips, according to the theory of probability, she should obtain 500 heads and 500 tails. But I'm sure the result will not be exactly 500 vs. 500. Therefore I think it's impossible to compare likelihood*.

For him, probability can be measured but likelihood cannot.

Summary

Chance cannot be measured or predicted mathematically leads to the conclusion that it is impossible to compare the likelihood of different outcomes. Underpinning this seems to be the idea that outcomes depend on luck. Although the misconception decreased with age, it still exists in a relatively large number of grade 12 students even after they have had formal training in probability. It seemed both context and data do not play a role in eliciting this misconception.

Group 5: Equiprobability

The literature reviewed in chapter 2 showed that equiprobability responses have been widely reported in other research and it was also one of main misconceptions observed in this study, but the analysis in this study is different from that in the literature. Suppose there are n possible outcomes of a trial, the equiprobability responses were subdivided into three categories: (1) All the n possible outcomes have a 50% chance; (2) each outcome has an equal chance 1/n and (3) if the chances of n possible outcomes are close to each other, they are the same in practice.

Examples

For some students, all possible outcomes of an experiment can only have two results, they occur or do not occur, each with a 50% chance. Item I4(4) asked students whether it is impossible, possible or certain that in drawing three cards, one from each pack, the result would be the 2 of diamonds, 5 of diamonds and the 8 of spades. Student f1061340, a grade 6 student in an advanced school, gave the following reply:

Answer: *Possible.*
Reason: *There are two possibilities: the three cards or not the three cards. I believe both of them have a 50% chance. I cannot say definitely that the outcome is impossible or certain, so I choose the possible option.*

Item II2(6) presented a situation where students were told that a mathematician predicted that he had a 30% chance of pulling out a white marble from a bag without looking, and then actually pulled out a white marble. They were asked whether they thought that the judgement was accurate or not. A grade 8 student, h5081441 replied:

Answer: *I think his prediction is inaccurate.*
Reason: *Only black marbles and white marbles are in the bag. He should have 50%, not 30% chance of picking out a white marble.*

A grade 12 student, g7124231, who has received instruction in probability, used equiprobability on two lucky draw items, III4(1) and III4(2). For the first item (200 girls vs.1000 boys), he wrote:

Answer: *(c) it is equally likely to pick out a boy's as a girl's name*
Reason: *No matter that there are more boys' names or more girls' names in it, the probability of being picked is equal.*

Then he was asked how about the situation of the 71st draw after 23 girls' and 47 boys' names were picked out. He still chose (c) and wrote:

Reason: *The condition you give is just a camouflage. The context of judging is the same as question (1). So I apply the equiprobability principle again.*

He believed in equiprobability and called it a principle, although there is no such principle in his textbook.

In many cases equiprobable responses were due to the belief that everyone is equal if a game is really fair. For example, in item III1(2) a grade 6 student, a3064020, when writing her reason for the equiprobable answer stated that it was because *picking out a boy or a girl is decided by God.* Lecoutre (1992) called such reasoning as using a <u>chance model</u>. It is equiprobable because it is by chance.

In all the above examples students distributed chance equally between two events that were compared. These types of example are similar to problems that have been used in other research. In this study the use of items with more than two possible outcomes items (category IV items) enabled the researcher to explore possible extensions of this idea. These types of item have rarely been mentioned in other research.

When faced with items that involve more than two possible outcomes, for example, three outcomes, some students still thought that each of the three had a 50% chance of happening, but some students gave slightly more complex answers. They thought that each outcome should have an equal chance of 1/3. For example, item IV5 involved two bags that containing red marbles and black marble. The students were asked to select the correct statement from the following choices:

a) it is most likely that both marbles are red
b) it is most likely that both marbles are black
c) it is most likely that one marble is red and the other one is black
d) it is impossible to indicate which one is the most likely among the three outcomes

A grade 12 student, i6122461 gave the following answer and reason:

Answer: *No option can be selected.*
Reason: *The likelihood should be equal.*

His classmate, f6122290m, also expressed the same thinking in answering item II1(9) when she wrote that *there are only three possibilities for a game: win, draw or lose. Each has a 33% chance.*

Another kind of thinking that also led to equiprobable responses was that close chances are the same in practice. Consider the following example given by a grade 8 student, f5081341, in an advanced school. He explained his equiprobable answer to item III3(3) in this way:

> *After the first 70 draws, there are 770 names left in the box. 365 girls' and 405 boys' names. The probability for girls is 47% and for boys is 53%. That is, if you draw 100 times, 47 times a girl's name will be picked out and 53 times a boy's name will be picked out. The difference is not very big. But if you only draw once, boys and girls should be equal.*

Frequency

Although some equiprobable responses were observed in items from categories I and II, they were mainly observed in items from categories III and IV. In fact only 10 students used this reasoning just in category I or II. As was done in Group 4, the following data analysis focused on the 30 chance comparison items in categories III and IV. The results are shown in Table 4.7.

Table 4.7 Number of students with the equiprobability misconception on the 30 chance comparison items

Total number of students in each stream for each grade		Grade 6*		Grade 8*		Grade 12*	
		Adv	Ord	Adv	Ord	Adv	Ord
		50	124	101	108	97	87
Number of students who used the misconception	0 out of 2; 0 out of 3; 0 out of 4	43(86)	96(77)	76(75)	94(86)	81(84)	70(81)
	1 out of 3 or 1 out of 4	3(6)	15(12)	9(9)	5(5)	5(5)	8(9)
	1 out of 2 or 2 out of 4	1(2)	6(5)	3(3)	3(3)	6(6)	7(8)
	2 out of 3 or 3 out of 4	2(4)	6(5)	11(11)	3(3)	5(5)	1(1)
	2 out of 2; 3 out of 3; 4 out of 4	1(2)	1(1)	2(2)	3(3)	0(0)	1(1)

* Figures in brackets represent the percentage of students using the misconception for the corresponding cohort.

An examination of the data in Table 4.7 indicates that about 20% of the students in each grade and each stream used equiprobability at least once, with about half of these students using the misconception at least half of the time. This means variables such as school type, students' age, background in probability did not effect the use of the equiprobability approach. Compared to other research, it seems that equiprobable responses were not as common for Chinese students as for students in the West. The possible explanations will be given later in the section. It seems that the different results might depend mainly on the form of the item.

Data and Context

An examination of Table 4.8 shows some interesting patterns.

Table 4.8 Location Index for 30 chance comparison items in eliciting equiprobability responses

Item	Location		Item	Location		Item	Location
IV9	*		III9	-0.731		III2(4)	0.889
IV2	-2.442		IV10	-0.730		III2(1)	0.913
IV7	-1.390		III4(1)	-0.353		III4(2)	0.970
IV6	-1.390		III5	-0.258		III1(3)	1.053
III10	-1.282		IV5	-0.153		III1(4)	1.053
IV8	-1.145		III3(4)	0.355		III1(2)	1.383
IV4	-1.119		III7	0.447		III2(3)	1.383
III8	-1.027		III4(3)	0.593		III3(1)	1.680
III2(2)	-0.920		III4(4)	0.593		III3(3)	1.858
III6	-0.874		III3(2)	0.644		III1(1)	*

* Item IV9 and III1(1) were deleted as extreme items by the computer when running the analysis.

First, it is far more likely that students will use an equiprobability argument in category III than category IV items. That is, one-stage items are more likely to elicit the misconception than two-stage items. Except for IV5 and IV10 all 8 items in category IV are located in the lowest 10 positions, the positions indicating that they are least likely to elicit an equiprobability response. Secondly, equiprobable responses were observed more in lucky draw items (involving drawing names out of a box) than marble items or spinner items. Thirteen lucky draw items occupied the highest 14 locations. Thirdly, for the lucky draw context, it seems that items involving close data, items III1 and III3 (girls and boys numbers were 20 vs. 22, 400 vs. 440) elicit more equiprobable responses than items III4 (200 girls vs. 1000 boys) where the data were far apart. Finally, the first draw items (items (1)) or the next draw after equal number of boys and girls were picked out items (items (3)) seemed more likely to elicit the misconception. There is no clear pattern in the marble and spinner data.

Evidence that data effected the use of equiprobability was also obtained from the interviews. A grade 6 student, a2063021m, used equiprobability in answering item III1(1) where the number of girls and boys was very close (20 vs. 22). He said, *only two more so the same chance*. When he was asked what he would conclude if the numbers were changed to 20 girls 40 boys, he said, *if so I believe the boys have more chance than the girls have*.

It is particularly interesting to investigate how grade 12 students who had some knowledge of probability explained their equiprobable responses. Student a6122061, was a grade 12 student in an advanced school. He answered the same item, III1(1) with the following reason,

> Reason: *As the name to be picked out only has two possibilities a boy or a girl, P(A)=1/*C_2^1.

Another grade 12 student, b5121071m, explained his equiprobable response further in the interview. In the questionnaire his original answers for items III1(3) and III1(4) were correct, but they were crossed out when he made the final choices. He changed both of them to equiprobable answers and reasons. In the interview, when he was asked to explain the changes in the questionnaire, he said

> *It is because only one name is picked. Only after a long run would the ratio come true and the difference of likelihood could be identified…. I think after at least several hundred trials, the difference could be identified.*

The conclusion that data and context play a role in eliciting equiprobable responses is consistent with other research (Green, 1982; Lecoutre, 1992; Williams & Amir, 1995).

Other

It has been mentioned that compared to other research it seems that equiprobable responses were observed less often in this study. For example, an average of 42% of the students (aged 11-16) in Green study (1982) exhibited this type of reasoning in solving the following item:

> A mathematics class has 13 boys and 16 girls in it. Each pupil's name is written on a slip of paper. All the slips are put in a hat. The teacher picks out one slip without looking. Tick the correct sentence:
> (A) The name is more likely to be a boy than a girl
> (B) The name is more likely to be a girl than a boy
> (C) It is just as likely to be a girl as a boy
> (D) Don't know

In this study, 28% (13 out of 47) of the grade 6 and 8 students (aged 11-13) applied the equiprobability approach in answering item III1(1) (this item had the highest location in eliciting equiprobable responses), an item which is similar to the Green item. The conditions for the two items are similar, 13 boys and 16 girls vs. 22 boys and 20 girls. But in Green's study, students were not required to give their reason. The main difference is in option (d). There is a difference between don't know and it is impossible to compare the likelihood of the two outcomes. In Green's study students who were confident in their ability but unclear how to make the comparison might have preferred the equiprobable option rather than admit that they did not know. Actually, about 34% (16 out of 47) grade 6 and 8 students in this study selected the option that it is impossible to make the comparison, while only about 2% of the students in Green's study selected the don't know option. This explanation, together with findings such as the use of this misconception is data and context dependent means that variability between research studies might be expected.

Another possibility is that incomplete or unclear answers in this study were coded as not using the misconception and, consequently, the number of coded responses is less than the actual number. For example, in reading lucky draw items, some students took *pick out the seventh slip* to mean *pick out seven slips* and *pick out the seventy-first slip* to mean *pick out seventy-one slips*. In the interview, some of these students changed to a correct answer, but some of them applied the equiprobability misconception. In the questionnaire student f6122290m gave the equiprobable answer to item II1(9) mentioned earlier but chose option (d) (impossible to make a compare) for items III3(3) and III3(4), where she was asked about the 71st slip. Looking at her written answer it was unclear if she had misread the problem. In the interview she used equiprobability consistently, no matter how the numbers were changed.

Summary

Equiprobability was one of the main misconceptions observed in this study, although it was used less frequently than reported in other research. It was also found that data and context impacted on the use of this misconception. This research separated equiprobability responses into three sets: each outcome with 50% chance; all outcomes equally share an equal chance; and close chances are the same in practice. Although the data on students who used only one version of equiprobability is available from the main study it is not reported. Since an equiprobable option was not included in the multiple choice alternatives for three possible outcomes items (Category IV) it was sometimes difficult to classify whether a student chose it is impossible to indicate which one is the most likely among the three outcomes because of equiprobability or another reason. However, the data obtained in teaching experiment that is reported in Chapter 6, where the equiprobable option was added and all the students were interviewed, will address this question.

Group 6: Outcome approach

This group consists of two misconceptions that are labelled as the outcome approach and the weak outcome approach. The outcome approach label is based on Konold's (1989) concept and the weak outcome approach is an extension of his concept. According to the description given by Konold (1989), students who use the outcome approach usually use the chance of an event, more than or less than 50%, as a guide to determine the certainty of an event happening. Frequency information is not used. For example, if the chance is greater than 50%, students using this misconception believe the outcome will certainly happen, otherwise, it will certainly not happen. If the actual result is unexpected, there must be a reason. If you predict that an outcome has a chance of occurring close to 100% and it actually happens they are certain that the prediction is accurate, otherwise, it is inaccurate. The studies of Konold (1989, 1991 and 1995), Green (1982), Fischbein et al. (1991) and Williams and Amir (1995) concluded that the outcome approach is used by adolescents and college students.

Examples

The following are a few sample responses from students illustrating how the outcome approach was used. Item II3(1) asked students to select a choice from the following that best illustrated the meaning of tomorrow there is an 80% chance that it will rain.

a) it will certainly rain tomorrow
b) it will certainly not rain tomorrow.
c) suppose that there were 10 days in a year the forecaster said that "tomorrow there is an 80% chance that it will rain", and on around 8 out of the10 days it rains the next day
d) suppose that there were 10 days in a year the forecaster said that "tomorrow there is an 80% chance that it will rain", and on exactly 8 out of the10 days it rains the next day

A grade 6 student, g1061381, gave the following answer and reason:

Answer: *(a)*
Reason: *Because 80% is very close to 100%, and usually when over 50%, it means it will certainly rain.*

Besides taking 50% as a guide, some other students in this study used a parallel but a bit sophisticated criteria, that is, comparing the likelihood that an outcome happens with it not happening.

For the same item, a grade 8 student, g4083400, gave the following answer and reason:

Answer: *(a)*
Reason: *Because he said the chance of rain was 80%. It means 80% of rain and 20% of no rain. 80% is higher than 20%. So it will certainly rain tomorrow.*

Item II2(5) presented a situation where they were told that a mathematician predicted that he had a 50% chance of pulling out a white marble from a bag without looking, and then actually pulled out a white marble. They were asked whether they thought that the prediction was accurate or not. A grade 6 student, e2062220 replied:

Answer: *Not very accurate.*
Reason: *As he said the chance was only half, and actually the marble he picked out was white. If he had predicted his chance was 90%, I would say his prediction is accurate.*

Item II1(7) presented a situation where a manager predicted his team had an 80% chance of winning and students were asked which of the following choices could be considered as an indication that the prediction was very accurate.

a) his team really wins this match
b) his team really loses this match
c) suppose that the match could be repeated 10 times, his team wins all the 10 matches
d) suppose that the match could be repeated 10 times, his team wins 9 out of the 10 matches
e) suppose that the match could be repeated 10 times, his team wins 8 out of the 10 matches

A grade 12 student, i5121460 gave the following answer and reason:

Answer: *(a)*
Reason: *The manager's prediction means the chance of winning is much higher than that of losing. So his team should win the match.*

These students appear to judge the accuracy of a prediction after one trial or to decide whether an event will occur or not occur after comparing its chance value to 50% or the chance value of its complement.

It is interesting to consider how the outcome-oriented students explained a 50% chance value. Some of them interpreted the middle value as a "total lack of knowledge about the outcome" (Konold, 1995). A grade 6 student, a2063011m, said in his interview:

> *If a weather forecaster said 50% chance of rain in 100 out 365, nearly 1/3 days in a year, I think his judgement is wavering and unbelievable. … Yes, it means he is not a good forecaster.*

Other outcome-oriented students believed that 50% means neutral, both occurring or not occurring are acceptable outcomes. In an interview, when a grade 6 student, f2062240m, was asked to judge if a prediction involving 50% chance was accurate or not, she said that she could not make the decision, but she could make a judgement if the percentage was changed to 80%.

There is a particular manifestation of the outcome approach that this researcher will refer to as the **firm outcome approach**. These students form a subset of those students who were labelled as using the outcome approach. They preferred to use the outcome approach in each trial. For example, in answering item II1(7), which has been mentioned previously, a grade 12 student, i7124290, answered this way:

> Answer: *(c) suppose that the match could be repeated 10 times, his team wins all the 10 matches*
> Reason: *His team has 80% chance for each match. That is, the probability of winning is much higher than that of losing. In this case the manager's prediction is proved to be true every time.*

Some other students applied the outcome approach, but with a small adjustment. As the following example shows, neither interpreting an 80% chance of rain to mean that it would certainly rain or using the formal idea of frequency, they preferred to say that it would probably rain. Since there was no such option in the multiple-choice answers, they chose 9 out of 10 such days it rains rather than the option that it will certainly rain or that on 8 out of 10 such days it rains. For item II3(7), a grade 8 student, h1082371m, elaborated as follows in the interview:

> *For me, 9 days is the answer that fits. I don't agree with definitely rain or not rain. Neither with 10 out of 10 such days because it conflicts with 80%. Eight days is a bit less. Nine days is the best answer.*

These students cannot be labelled as using an outcome approach, as they do not agree with the conclusion that would usually be drawn by a student who used the outcome approach as defined by Konold. They can be considered to be leaning towards such an approach and, consequently, such a misconception is called a **weak outcome approach**.

Frequency

Note that since the firm outcome approach is a subset of the outcome approach, the following tables include both students who were labelled as having an outcome or firm outcome approach. However, since the weak outcome approach is an extension of Konold's concept, weak outcome responses have not been included (15 students in total).

The outcome approach was one of the most common misconceptions observed in this study. Although items in category I could elicit this approach, items in category II were designed to focus on this misconception. In fact, there were only 15 students who used this approach in category I and did not use it in category II. Consequently, the following data only focus on category II. Table 4.9 contains the number of students who used the outcome approach.

Table 4.9 Number of students with the outcome approach misconception on the 27 Category II items

Total number of students in each stream for each grade		Grade 6*		Grade 8*		Grade 12*	
		Adv	Ord	Adv	Ord	Adv	Ord
		50	124	101	108	97	87
Number of students who used the misconception	0 out of 3 using outcome approach	30(60)	70(57)	59(59)	68(62)	68(70)	59(68)
	1 out of 3 using outcome approach	15(30)	35(28)	26(26)	33(31)	26(27)	22(25)
	2 out of 3 using outcome approach	4(8)	15(12)	13(13)	6(6)	3(3)	4(5)
	3 out of 3 using outcome approach	1(2)	4(3)	3(3)	1(1)	0(0)	2(2)

* Figures in brackets represent the percentage of students using the misconception for the corresponding cohort.

The data show that the outcome approach is a common misconception. About 40% of grades 6 and 8 students and about 30% of grade 12 students used it at least once. More than 10% of grade 6 students (both advanced and ordinary) used it in at least two out of three cases. This is similar to the figure for the advanced grade 8 classes, but is considerably lower for all the other groups. There is no pattern that clearly differentiates the advanced and ordinary students in their use of the outcome approach.

There were comparatively few responses that were classified as the firm or weak outcome approach. However, it seems that students who used the firm or weak outcome approach were more likely to use the outcome approach in other items. Ten out of 15 students with the weak outcome approach and five out of 11 students with the firm outcome approach applied the outcome approach to other category II items.

Data and Context

An examination of Table 4.10 shows some interesting patterns. There were nine items involving a mathematician [2(n)], nine parallel items involving a coach [1(n)] and nine parallel items involving a weather forecaster [3(n)]. Of the 9 items involving a mathematician seven were in the lowest eight locations in terms of eliciting the outcome approach. It would appear that, except for 2(4) and 2(5), this context is less likely to result in students using the outcome approach. In comparison, for the nine forecaster items, seven were located in the top half of items eliciting the misconception, the exceptions being except 3(3) and 3(8). It appears that weather forecaster items are more likely to elicit an outcome approach response. The coach items are spread between the middle and top third locations. Of the five items most likely to elicit the misconceptions, four were coach items. It would seem reasonable to conclude that the context does play a role in eliciting this response.

Table 4.10 Location Index for 27 items in eliciting outcome approach responses

Item	Location		Item	Location		Item	Location
2(8)	*		1(8)	-0.434		3(7)	0.685
2(9)	-1.719		1(3)	-0.077		2(4)	0.703
2(7)	-1.627		2(5)	-0.031		1(7)	0.825
2(2)	-1.143		1(9)	0.081		3(4)	0.892
3(3)	-1.106		1(2)	0.154		1(5)	0.998
2(6)	-0.808		3(1)	0.401		1(6)	1.113
2(1)	-0.657		3(2)	0.445		1(4)	1.220
2(3)	-0.636		3(6)	0.449			
3(8)	-0.513		3(9)	0.584			
1(1)	-0.475		3(5)	0.675			

* Item 2(8) was deleted as an extreme item by the computer when running the analysis.

Seven out of the ten items most likely to elicit an outcome approach, used an open-ended format (item (4), (5) and (6)). So in addition to the context, per se, format plays a role in eliciting this response.

However, it seems that the role of data does not result in any clear pattern. For example, items involving a 30% chance ((3), (6) and (9)) are spread all over the location table. There are five items with the 80% chance ((1), (4) and (7)) in the top third, but the others are spread widely. The conclusions from these analyses are that: (1) Mathematician items were less likely to elicit the outcome approach than items involving the other two contexts; (2) Open-ended items were more likely to elicit the outcome approach than multiple choice format items; and (3) There is no clear pattern relating to the role that data played in eliciting the outcome approach.

However, the role of data (especially a 50% chance) cannot be completely ignored. An example of the effect that data can have was obtained during an interview. In answering item II2(8), which presented students with the situation where there was a 50% chance of pulling out a white marble, a grade 12 student, c5121131m, said:

> *I think it does not matter if a black or white marble is pulled out, both the situations indicate his prediction is accurate. But if you change the percentage from 50% to 80%, I will say his prediction is very accurate when he pulls a white marble out 80 times and a black marble 20 times in 100 trials.*

When he was asked why he used different strategies in answering the two items, he said:

> *50% means perhaps a white, perhaps a black. But 80% has a different meaning. The latter strategy does not fit solving a 50% chance problem.*

His understanding of the meaning of 50% chance is similar to the grade 6 student, f2062240m, mentioned earlier.

Other

In this study, blank answers, incomplete or unclear answers were coded as not using misconceptions. So the actual number of students who used either the outcome approach or the weak outcome approach may be more than the number indicated. For example, consider student a4083010m, a grade 8 student from an ordinary school. She had difficulty giving a written reason. She chose the weak outcome option in item II3(8) but left the explanation blank. The interview provided data that showed she was actually using the weak outcome approach.

Deciding whether or not a student used the outcome approach in open-ended items was more difficult. In the interviews, students were often asked to answer parallel items where the percentages or the results were changed. Some of them appeared to use the outcome approach, while others did not.

Item II3(6) posed a situation that a weather forecaster predicted that tomorrow there was a 30% chance of rain but actually it did not rain the next day. The students were asked whether they thought that the prediction was accurate or not. A grade 6 student, i2063360m, gave the following written reply:

> Answer: *Accurate.*
> Reason: *Since he doesn't say 100% chance it will rain tomorrow so it means the chance of no rain exists.*

It is not clear how she concluded that the prediction was accurate. In the interview, she was asked to answer a parallel item.

> Interviewer: *Suppose he predicted there was an 80% not a 30% chance of rain but actually it didn't rain the next day. Does it influence your decision?*
> Student: *Yes, I will say his prediction is inaccurate.*

She changed her decision without any hesitation. The data implies that she was actually using the outcome approach. However, some other students did not make any change to their decision when new parallel questions were asked and they explained it as both the target event happening or not happening were possible, so they could not say the prediction was inaccurate. Clearly, they were not using the outcome approach.

Summary

The discussion of the misconceptions in this section lead to the following major conclusions: (1) The outcome approach is a common misconception among Chinese students; (2) There is an extension of the outcome approach, the weak outcome approach that seems to exist; (3) The outcome approach is most common among grade 6 and advanced grade 8 students; and (4) The use of the outcome approach is dependent on both context and the format used to present the problem.

Group 7: One trial is unrelated to other trials

In addition to the outcome approach, one trial is unrelated to other trials was another misconception that led the students to interpret probability in a non-frequentist way.

Examples

Items in category II were designed to be parallel. One variation was changing the data from 80% to 50% to 30%. In the items one or two of the options were one trial options and the others were repetition options. Since the student with this misconception believed that one trial is unrelated to other trials, they appear to have had only three realistic choices: select one option that involves one trial, select one option that involves the least repetitions, or indicate that none of options could be selected. Here are three examples.

Item II3(7) posed the situation that tomorrow there is an 80% chance that it will rain and students were asked which of the following options could be considered as an indication that the prediction was very accurate.

a) it really rains the next day
b) it really doesn't rain the next day
c) suppose that there were 10 days in a year the forecaster said that "tomorrow there is an 80% chance that it will rain", and on all the 10 days it rains the next day
d) suppose that there were 10 days in a year the forecaster said that "tomorrow there is an 80% chance that it will rain", and on 9 out of the 10 days it rains the next day
e) suppose that there were 10 days in a year the forecaster said that "tomorrow there is an 80% chance that it will rain", and on 8 out of the 10 days it rains the next day

A grade 6 student, h2063320, gave the following answer and reason:

Answer: *(a)*
Reason: *It is talking about the weather on one day, not 10 days. 80% chance of rain means that rain is more likely. So I choose option (a) that it really rains the next day.*

She made a choice between two one-trial options.

Another grade 6 student from an advanced school, b1061111, chose the least repetition option for item II1(8). The item was parallel to the above item but the context was a football match and the data was 50%. He was asked to make a choice from the following options:

a) his team really wins this match
b) his team really loses this match
c) suppose that the match could be repeated 10 times, his team wins 5 out of the 10 matches
d) suppose that the match could be repeated 10 times, his team wins 3 out of the 10 matches
e) suppose that the match could be repeated 100 times, his team wins 50 out of the 100 matches

The student chose (c) and gave the following reason:

Among the five options, only (c) and (e) are close to the meaning of the coach's prediction. Since the coach indicated this match, comparatively, only option (c) is the closest answer.

Students who thought both the outcome approach option and frequency approach option were wrong gave answers similar to the following. In answering item II3(8), a grade 8 student, a5081051, wrote

Answer: *None of the options can be selected.*
Reason: *In my opinion, all the options are wrong. Because 50% only refers to tomorrow. It means only that day. It is not related to 10 days or 100 days. So options (c), (d) and (e) are excluded. And options (a) and (b) also don't mean the weather forecaster's prediction is very accurate.*

Frequency

Table 4.11 shows the misconception was not observed very often in grade 6 students. Overall, it seems that advanced students used it more often than ordinary students. The grade 12 students who had some limited formal exposure to probability used it more often than grade 6 students, who had not been taught probability. However, since the numbers for each grade were small no clear pattern was found.

Table 4.11 Number of students with the misconception of “one trial is unrelated to other trials” on the 27 Category II items

Total number of students in each stream of each grade	Grade 6		Grade 8		Grade 12	
	Adv	Ord	Adv	Ord	Adv	Ord
	50	124	101	108	97	87
Number of students with the misconception	3	3	9	4	7	5

Data and Context

Seven out of the 31 students used the misconception twice, so a total of 38 responses were observed using the misconception. In order to examine the effect of data and context, the 38 responses were categorised twice in Table 4.12.

Table 4.12 Number of “one trial is unrelated to other trials” responses classified by data and context

Data and context	Data				Context		
	80%	50%	30%		Coach	Mathematician	Forecaster
Number of responses	12	15	11		16	3	19

The role of context seemed much stronger than that of data. Students accepted the frequency interpretation of probability much easier in mathematician items in which the mathematician predicted his chance of pulling a white marble from a bag. Comparatively, coach items and weather forecaster items elicited more one trial is unrelated to other trials responses. One possible explanation is that it seemed the students thought the results of a match or the weather are greatly affected by many things that are out of peoples control, but the result of pulling out a marble from a bag is not. So it was much more difficult for a coach or a weather forecaster to make a prediction that tallies with the actual situation. Furthermore, repeating a match many times might be considered unrealistic, while repeating pulling a marble out of a bag could easily be done. The issue of the potential unrealistic nature of the coach task is discussed in Group 9. Data does not appear to be a factor in eliciting the misconception.

It should be noted that this misconception was not very common and in most cases it was only used once by a student, so any conclusions should be considered as very tentative.

Other

In the interviews some students explained why they did not choose any of the repetition options. A grade 8 student, i1082431m, in an advanced school answered item II1(7) in the interview:

Student:	*The coach is talking about* ***one game****. 80% chance of winning is quite higher, so I choose (a).*
Interviewer:	*How about the options (c), (d) and (e)?*
Student:	*I will not select them. As it is possible that his team wins or loses each game of the10 games.*

He chose the team really wins this match since the coach's prediction was for one game and not many games, and he believed anything could happen in repetition situations. When he answered item II2(2), he also argued that *it is possible that all the 10 or 100 marbles he takes out happen to be white or all of them happen to be black. … It is nonsense to do a lot of trials.* For this student no pattern existed even after a lot of repetitions. Such thinking is similar to the misconception that chance cannot be measured mathematically.

His schoolmate, a grade 6 student, f1061321m, also explained that considering the uncertainty of the results of the repetition of trials, he preferred a one-trial option.

The option that involves one trial was designed primarily to investigate the outcome approach. The data indicate that this misconception is related, in terms of student use, to both the misconception chance cannot be measured mathematically and the outcome approach. Actually about 50% and 40% of the students who thought that one trial is unrelated to other trials also used chance cannot be measured mathematically and the outcome approach in other items, respectively.

Summary

In addition to the outcome approach, the misconception that one trial is unrelated to other trials was another bias that meant the students did not interpret chance as a measure of how often an outcome would happen. It seems that context might play a role in eliciting this misconception. Since the number of students with this misconception is small this conclusion should be considered tentative. However, it is a misconception that would be worth further research.

Group 8: Interpreting chance by data matching or word matching

Some students realised that to interpret chance by the outcome approach was wrong, as the chance value given was not 0% or 100%. They preferred to interpret chance by using frequency. However, their explanation was still wrong because it was based on interpreting chance by data matching or word matching. Data matching means that students think that a 30% chance of an event happening means that it will happen exactly three out of

10 times, since 30% is matched with exactly three out of 10. For word matching they think that since a 30% chance and not *around* a 30% chance was used, one must select an answer that exactly matches the number 30%.

Examples

In order to examine students' understanding of uncertainty, options such as around 3 out of 10 and exactly 3 out of 10 were used together in chance interpretation items. For example, item II1(3) asked the students to make a choice from the following that best illustrated the meaning of the team has a 30% chance of winning.

a) his team will certainly win this match
b) his team will certainly lose this match
c) suppose that the match could be repeated 10 times, his team wins around 3 out of the 10 matches
d) suppose that the match could be repeated 10 times, his team wins exactly 3 out of the 10 matches

A grade 6 student, a1061041 gave the following answer and reason:

Answer: *(d)*
Reason: *Because 30% = 3/10, the answer should be (d).*

It seemed he chose exactly in order to match the data. A grade 8 student, e1082240, in her answer of item II1(1), expressed the same thinking. She preferred the exactly option rather than the around option because the around option *(c) is not so accurate.* These students appeared to think that if around were chosen, the frequency would not match the meaning.

Comparatively, fewer students made their choices based on word matching. They thought that since the word around did not appear in a prediction, the best explanation of the prediction should not involve the word. For example, a grade 8 student, a4084011m answered item II1(3) in this way:

Answer: *(d) suppose that the match could be repeated 10 times, his team wins exactly 3 out of the 10 matches*
Reason: *As what he said was 30%, not around 30%.*

A grade 6 student, h2063350, expressed the same idea in her answer to item II1(2) when she wrote that *around 50% was not mentioned in the question, so the word around should not be mentioned in the answer either.*

Two grade 8 students chose 100 experiments (the most repetitions option) because they took the percentage symbol as the basis for their answer. Item II3(8) asked the students to indicate the option that could be considered as an indication the prediction that tomorrow there is a 50% chance of rain was very accurate. A grade 8 student, a4084031, gave the following answer and reason:

Answer: *(e) suppose that there were 100 days in a year the forecaster said that "tomorrow there is a 50% chance that it will rain", and on 50 out of the 100 days it rains the next day*
Reason: *The chance value was measured by percentage, so I should answer 100 days to match it.*

Frequency

Table 4.13 Number of students with the misconception of "interpreting chance by data matching or word matching"

Total number of students in each stream of each grade	Grade 6* Adv 50	Grade 6* Ord 124	Grade 8* Adv 101	Grade 8* Ord 108	Grade 12* Adv 97	Grade 12* Ord 87
Number of students with the misconception	6 (12)	27 (22)	15 (15)	14 (13)	6 (6)	7 (8)

* Figures in brackets represent the percentage of students using the misconception for the corresponding cohort.

The number of students with the matching misconception is summarised in Table 4.13. An analysis of the data shows that a relatively large number of both grade 6 and 8 students used the misconception. The difference between the two streams was not very pronounced in grades 8 and 12.

Data and Context

Except for the two students who took the percentage symbol as the basis for their answer, all the other students used the matching misconception in one type of item where they were asked to select the option that had the closest meaning to the prediction (items II1(1)-(3), II2(1)-(3), II3(1)-(3)). So the analysis of the effect of data and context will focus on these nine items. Each questionnaire included one of the nine items, which meant that each student had one opportunity to apply the misconception.

Table 4.14 Number of students interpreting chance by data matching or word matching classified by data and context

Data and context	Data				Context		
	80%	50%	30%		Coach	Mathematician	Forecaster
Number of responses	32	13	28		22	38	13

It seemed that items with a 50% chance did not elicit as many matching responses as the items with an 80% or 30% chance. This result might not be due to data, per se, but the design of the options. The form of items with 80% and 30% ((1) and (3)) was exactly parallel, but the design of the items with 50% was different. Instead of including two certain options involving one trial (for example, will win/lose that game), one uncertain option involving one trial (for example, may or may not win that game) was used and an option involving 100 trials was added in the items with 50%. The may or may not option was selected by many students and this could have decreased the number using the matching misconception.

As for the role of context, many more matching responses were observed in mathematician items. Probably this result can be explained in a similar manner to that given in the last group, namely, that the mathematician context is a controlled situation that can be repeated easily.

Other

Most of the students with the matching misconception thought that since the probability value is a fixed ratio it should be interpreted precisely. A grade 12 student, e5121211m, expressed this idea by using an equation. He was given item II1(1) that asked the closest meaning of has an 80% chance to win a match.

Answer:	*(d) suppose that the match could be repeated 10 times, his team wins exactly 8 out of the 10 matches*
Reason:	*P(robability) = x/10 = 0.8, then x = 8.*

In the interview, he said that *the probability is equal to games won over total games. … As the probability is clearly given, I believe the word around is unnecessary.*

The problem he had was not in how to measure probability but in how to interpret probability. In the interviews, many younger students explained their choice of around as follows. They indicated that it was because the word prediction was used or they doubted the accuracy of the prediction.

Summary

The underlying thinking seems to be related to a deterministic view of the concept of probability. This view can be found in students' perspectives on other measurement concepts. For these students chance is treated in the same manner as the distance to a friend’s house (3.5 km) or the height of a person (1.64 m tall), where the uncertainty of the measurement is usually ignored. With this perspective they thought that a 30% chance of an event should be interpreted as it would happen exactly three out of 10 times.

Group 9: Increasing repetition is not better for predicting

Group 9 discusses the misconception that increasing repetition does not help in making a better prediction or determining the accuracy of a prediction. This seems to be based on the belief that although some repetition is useful it is unnecessary to repeat again and again, so students usually chose a smaller repetition option.

Examples

Six items involving both smaller and bigger number of repetition options were used in this study. Some students chose both options for their answer, and their responses indicated that they thought there was no difference between a smaller sampling and a bigger sampling, as illustrated in the following example.

Item II3(2) asked the students to select which of the following options best illustrated the meaning of tomorrow there is a 50% chance of rain.

a) it may or may not rain tomorrow. The forecaster doesn't really know what the result will be
b) suppose that there were 10 days in a year the weather forecaster said that "tomorrow there is a 50% chance that it will rain", and on around 5 out of the 10 days it rains the next day
c) suppose that there were 100 days in a year the forecaster said that "tomorrow there is a 50% chance that it will rain", and on exactly 50 out of the 100 days it rains the next day
d) suppose that there were 100 days in a year the forecaster said that "tomorrow there is a 50% chance that it will rain", and on around 50 out of the 100 days it rains the next day

A grade 8 student, d1082170 replied in this way:

Answer: *(b) and (d)*
Reason: *This is because the meaning of the sentence is that the possibility of rain and no rain is half-and-half and both option (b) and (d) mean that.*

She chose the two options because both of them represent the same ratio. Another 11 students in this study expressed a similar idea.

Other students with the misconception chose the option with a smaller repetition. For example, consider the answer given to II1(8) by a grade 12 student, b6122080. The item presented a situation where a coach predicted his team had a 50% chance of winning and students were asked which of the following choices could be considered as an indication that the prediction was very accurate.

a) his team really wins this match
b) his team really loses this match
c) suppose that the match could be repeated 10 times, his team wins 5 out of the 10 matches
d) suppose that the match could be repeated 10 times, his team wins 3 out of the 10 matches
e) suppose that the match could be repeated 100 times, his team wins 50 out of the 100 matches

The student gave the following answer and reason:

Answer: *(c)*
Reason: *Based on 10 games, we have had strong evidence to say that the team has a 50% chance to win. The result of 100 games can be deduced from the 10 games.*

Another grade 12 student, i7123541, expressed similar thinking. In answering item II2(2) he wrote that *you get the probability by repetition, but it is unnecessary to do too many repetitions.*

Not all the students explained why they thought the option with a smaller number of repetitions was better than the one with a bigger number of repetitions. Some students only mentioned that the fewer repetitions agreed with the chance value but did not indicate clearly why the more trials option was not better, as illustrated by the following answer to item II2(8). Item II2(8) was parallel to item II1(8) that has been mentioned above with the context changed to pulling out a marble from a bag. A grade 6 student, c3064151, chose option (c) involving 10 trials and not option (e) involving 100 trials. He wrote that *you pull out 5 white marbles in 10 turns means your chance of pulling out a white marble is 50 percent. So option (c) is correct.* Since the student did not explain why the larger sampling was not better, it is unclear what made him choose (c) not (e). In the interviews, some students explained their reason in detail and another two justifications were identified.

The first justification was that the more trials the easier it is to get what you want. So if you got what you want with fewer trials, it meant your prediction was more accurate. For example, a grade 8 student, c5081130m, gave a similar answer and reason to student c3064151 when answering the same item, II2(8). In the interview, when she was asked why the option of pulling out 50 white marbles in 100 turns was not selected, she explained that *I think you need less turns to achieve the result of 50%, it indicates a more accurate prediction.* This justification was also observed in a written answer of a grade 6 student, a1061060, when answering item II3(8).

The second justification was that the more trials, then the more errors. So if you carried out more trials, more errors would be accumulated. In an interview, a grade 12 student, h6122421m said he selected the option with the fewer repetitions as *the more games played, the more fluctuation arises.*

Frequency

Forty-one students used the misconception. Since nine sets of questionnaire were used in each class, and only six sets included the six items with both smaller and bigger number of repetition options, the actual frequency was 41 out of 377, i.e., about 11% of the students.

Table 4.15 Number of students with the misconception of “increasing repetition is not better for predicting” observed on 6 items

Total number of students in each stream of each grade	Grade 6* Adv 36	Grade 6* Ord 81	Grade 8* Adv 67	Grade 8* Ord 72	Grade 12* Adv 64	Grade 12* Ord 57
Number of students with it	5 (14)	13 (16)	6 (9)	6 (8)	7 (11)	4 (7)

* Figures in brackets represent the percentage of students using the misconception for the corresponding cohort.

From looking at Table 4.15 it seems that in each grade there is no major difference in performance between advanced and ordinary school students. Grade 6 students used the misconception slightly more often than the older students.

Data and Context

All the six items involved *50% chance* so the impact of changing the data is not relevant. The 41 responses with the misconception were categorised into three different contexts (see Table 4.16) and the data in this table show that there does not appear to be any context effect.

Table 4.16 Number of “increasing repetition is not better for predicting” responses classified by different contexts

Context	Coach	Mathematician	Forecaster
Number of responses with it	13	13	15

Other

Although the justification of the more trials the more errors was not observed in written answers, several students referred to it in the interviews.

In answering an open-ended item II1(6), a grade 8 student, d5081181m, indicated that he could not judge whether the coach's prediction was accurate or not based on the result of one match. He suggested that information from 10 matches was needed. When he was asked how about repeating the match 100 times then making his decision, he said

> *It's not necessary, as it would lead a bigger deviation to the real value. 30% chance of wining means you win 3 out of 10 matches, about 30 out of 100 matches and less than 300 out of 1000 matches.*

Another example was given by a grade 12 student, g6122351m, in answering item II2(9). The item presented a situation where a mathematician predicted his chance of pulling out a white marble from a bag was 30% and the student was asked which of the following choices could be considered as an indication that the prediction was very accurate.

a) pulling out a marble and it happens to be a white marble
b) pulling out a marble and it happens to be a black marble
c) suppose that the game is repeated 10 times, black marbles are pulled out 10 times
d) suppose that the game is repeated 10 times, white marbles are pulled out 1 time
e) suppose that the game is repeated 10 times, white marbles are pulled out 3 times

He chose (e) and with the reason that *the probability is 30%.* When he was asked what he would do if a new option (f) suppose that the game is repeated 100 times and white marbles are pulled out 30 times was added, he indicated that he would not choose the new option. He said,

> *No, I still choose (e), 3 out of 10 times. Option (f) should be given like this: white marbles are pulled out about 30 times. Because with the increase of trials, the fluctuation should behave more obviously.*

Based on the information from the interviews, the real frequency of this misconception is higher than the data reported in Table 4.15.

Another 12 responses, which indicated that the larger number of repetitions was not better, were not labelled as using the misconception. This is because all the answers were observed in coach items and their **main reason** was that *it's unreal to repeat a match so many times*. Based on this, some of them selected an option with no repetition, some of them selected one with fewer repetitions and others said none of options could be selected. Since the conclusions were based on their rejection of the supposition that the game could be repeated, not by any misconception of probability, these responses were not counted in frequency tables.

A final note on this misconception should be added. It could be argued that repeating a match 100 or even 10 times is not realistic, and in addition, twelve responses clearly indicated that repetition was unrealistic. However, this would not apply to pulling a marble out of a bag 10 or 100 times or looking at the weather over 10 or 100 days. The researcher believes that this is a real misconception and is not based on any belief that the situation is unrealistic for some students. This is reinforced by the fact that this type of reasoning occurred equally for all three contexts.

Summary

There are students who interpreted probability by frequency, but their understanding is faulty. They believed that increasing repetition of an experiment does not result in being able to make a better estimate of probability. Also, this misconception does not seem to be context dependent, even when one of the situations could be considered unrealistic. The question of the role that data play is worthy of further research.

Group 10: Positive and negative recency

Some students think that what will happen in the future depends on what has happened in the past. This leads to two opposite kinds of responses. One is that after a long run of one outcome the bias will rectify itself in the future. The other is that the bias that appears in the past will continue in the future. These two kinds of responses have been studied in other research. The first kind type of response has been referred to as negative recency, or the gamblers' fallacy, and the second as positive recency. In this study, the question of how students respond in a situation where there is no bias in the past is also studied.

Examples

In a two-outcome experiment, students using negative recency believed that after a long run of one outcome, the other outcome was more likely to occur. The following example of using the negative recency was given by a grade 12 student in an advanced school, e6122241. In item III2(4), where students were asked how about the 7th slip after six boys' names were picked out in a class with 5 girls and 27 boys, he selected the option that the girls had more chance in the next turn, because

> *After mixing up thoroughly, the ratio between the number of boys' names and girls' names should be about 5.4 : 1 in every unit. Since all the 6 names are boys' names, certainly, it is more likely to pick out a girl's name in the 7th draw.*

The condition **mixed up thoroughly** was usually emphasised by these students.

Students using the positive recency believed that future results should follow the pattern that occurred in the past. Consider a parallel item III1(4), where students were asked how about the 7th slip after six boys' names were picked out in a class with 20 girls and 22 boys. A grade 8 student, b5081100, gave the following answer and reason:

> Answer: *(a) it is more likely to pick out a boy's than a girl's name this time*
> Reason: *As he has already continually picked out 6 boys' names, it means that picking out a boy's name is more likely. So the next slip should still be a boy's name.*

In addition to the situation where there was a run of a specific outcome, the situation where there was no bias in the past was also studied. Four items similar to III2(3) involve this condition. Item III2(3) posed a situation where the students were asked how about the 7th slip after three girls' and three boys' names were picked out from a class with 5 girls and 27 boys. A grade 6 student, d1061231 gave the following answer and reason:

> Answer: (c) *it is equally likely to pick out a boy's as a girl's name this time*
> Reason: *Last time the teacher picked out 3 boys' and 3 girls' names, equally likely. Although only 2 girls and 24 boys left, however, based on the result of the last time, I think boys and girls have the same chance.*

He and a few other students believed that the balance that appeared in the past should remain in future experiment. Since there is a common theme underlying both this reasoning and positive recency, these responses were also labelled as using positive recency.

Frequency

The number of students with the misconception is shown in Table 4.17. It was found that this misconception was not used very often in each grade and each stream. Negative recency responses were observed a little more often than positive recency responses.

Table 4.17 Number of students with the negative or positive recency misconception on 12 items

Total number of students in each stream of each grade	Grade 6		Grade 8		Grade 12	
	Adv 50	Ord 124	Adv 101	Ord 108	Adv 97	Ord 87
Negative recency	4	7	1	3	2	3
Positive recency (extended)	1	3	4	2	3	0

Data and Context

Sixteen out of the 20 students who used negative recency applied it under the condition that the lucky draw was undertaken in a class/school with close or equal number of boys and girls (Item III1(n) and III3(n)). This meant that negative recency is elicited more easily under this condition. Having close or equal numbers of boys and girls might make it easier for students to believe that the other outcome would occur in order to achieve a balance.

There was a different pattern for positive recency. Ten out of the 13 students used positive recency after a long series of draws (70 draws, Items III3(n) and III4(n)). It would appear that this situation made some students believe the bias that occurred in the past should be continue as it was *proved* again and again.

However, it is important to note that the small number of responses means that these conclusions should be considered tentative.

Other

Some students thought it was impossible to compare the likelihood of the first draw but the comparison could be made after a run of experiments. Here is an example. A grade 6 student, h2063310m, was given items III3(1) and III3(2) in the questionnaire. Item III3(1) asked her to compare the likelihood in the first draw undertaken in a school with 400 girls and 440 boys. Item III3(2) asked her how about the 71st draw after 15 girls' and 55 boys' names were picked out. For the first item she gave the following reply:

Answer: *(d) it is impossible to compare the likelihood of the two outcomes*
Reason: *Although boys are 40 more, in my opinion, it is uncertain that the next slip will be a boy's name. So I choose (d).*

For the second item, she believed the boys had more chance this time. In the interview, she gave her reason again for the first item.

Student: *It is impossible to make the comparison. Wholly depends on luck.*
Interviewer: *Could you tell me why do you think it is impossible to make the comparison in the first item but you believe the boys have more chance in the second item?*
Student: *Because you said the 70 names are 15 girls' and 55 boys', 55 is bigger than15, so a boy's name is more likely in the next draw.*
Interviewer: *If I change the 70 results to 55 girls' and 15 boys' names, will it influence your decision?*
Student: *Yes, I will say the girls have more chance.*
Interviewer: *If I change the 70 results once more, that they are all 70 boys' names. Will it influence your decision?*
Student: *I will say the boys have more chance.*

All the questions were about the likelihood of single slip, however, the student used the misconception of chance cannot be measured mathematically only in the first draw item but used positive recency consistently when there was previous information.

Other students gave a correct answer to the first draw item but used the two fallacies after being given the results of previous draws. Consider a grade 12 student, a5121011m, in an advanced school. He was given items III1(1) and III1(2) where the lucky draw was undertaken in a class of 20 girls and 22 boys. His written answers show that he applied different strategies for the two items. In the interview for the first item he said that the boys have more chance *as the chance is 22/42, higher than the chance for the girls*. For the second item, instead of calculating the chance, he gave the following reason and concluded that girls have more chance:

> *After 2 girls' and 4 boys' names are picked out, each has 18 names left. It should say they have the same chance. However, generally, I feel the ratio between boys and girls should be 11:10 at the beginning. But you say 2 girls and 4 boys, so the ratio is 2:1, bigger than the normal value. Therefore I choose option (b) to balance the ratio between the boys' and girls' number.*

This example indicated again that although some students know how to calculate probability in simple cases, their theoretical solutions often are challenged by their intuitions. The fact that a student can give an accurate chance value does not mean the student really trusts it. The misconception that sampling is a self-correcting process seems to prevail.

Summary

The positive and negative recency responses were not observed very often in each grade and each stream. It seems that under the condition that each of the outcomes has close or equal chance values, negative recency was more likely to be applied. Positive recency was more likely to occur after a long series of experiments. There is also some evidence to support the conclusion that even when students are able to calculate chance values their intuition prevails.

Group 11: Used own methods in chance comparison

All the students in grades 6 and 8 had never been taught probability in school. The students in grade 12 had been taught a short unit of about 10 lessons on the classical definition of probability from an equally-likely sample space perspective. For all grades, including grade 12 where they had had some exposure to calculating probability, some students used their own methods in chance comparison items. These methods seemed to be based on their real life experience, namely, that the greater the amount, then the greater chance. Group 11 will discuss these methods.

Both one-stage and two-stage chance comparison items were included in this study, and these situations are discussed separately. The first part relates to three kinds of methods used in solving one-stage items. They were: (1) the one with the greater absolute amount (greater number of black marbles in the different bags or larger area of blue on the different spinners) has more chance; (2) the one with the greater or lesser total amount (marbles or area) has more chance; and (3) the one with the greater or lesser difference has more chance. These three methods have been reported often in the literature which was reviewed in Chapter 2. For two-stage items there were different methods. They were: (1) the compound outcome, which is the combination of the more likely outcome at each stage has more chance; (2) the compound outcome associated with a greater total amount (marbles or area) has more chance; (3) if one amount is much more in one bag/spinner and, at the same time, it is not less in the other bag/spinner, this amount determines the compound outcome; and (4) the compound outcome that involves different results has more chance. These methods have rarely been reported in literature although are briefly mentioned by researchers such as Piaget and Inhelder (1975).

Examples

The easiest type of chance comparison tasks in this study was the lucky draw items. Based on comparing the number of boys and girls, students could get the correct answers. However, without applying the concept of ratio, this basic strategy could not be extended to complex tasks such as determining the bag that had the greatest chance of eliciting the target outcome. The following examples described how the students tried to extend the basic strategy to solve the items with two bags or two spinners without introducing ratio.

The first method was to choose the bag or spinner with the greater absolute amount that was related to the target outcome. For example, item III5 posed a situation where there were two bags. Bag A contained 8 red marbles and 16 black marbles, bag B contained 50 red marbles and 70 black marbles. The students were asked to determine from which bag it is more likely to pick out a black marble. A grade 6 student, e2063161m, preferred to choose bag B *as bag B has more black marbles.* Then he was given item III9 involving two different sized spinners. The bigger spinner B was an enlargement of the small spinner A. He applied the same strategy again and chose spinner B because *spinner B has more blue part.*

The second method was to choose the one with the greater or lesser total (number of marbles or area of spinner). Almost all the students with this bias in this study chose the one with the lesser total. In answering the item III6 involving bag A (21 red and 8 black marbles) and bag B (210 red and 80 black marbles), a grade 6 student, f2062260, believed there was more chance of pulling out a black marble from bag A. The reason she gave was:

> *This is because bag A only has 29 marbles, while bag B has 290 marbles, 10 times that of bag A, so it is much more difficult to pull out what you want.*

The idea that the greater amount the more difficult it was to pick out what you want was described mathematically by a grade 12 student, e6122250. The item posed to her was item III5 involving bag A (8 red and 16 black marbles) and bag B (50 red and 70 black marbles). She thought there was a greater chance of getting a black marble from bag A because

$$P(A)=1/C_{24}^{1},\ \ P(B)=1/C_{120}^{1},\ \ P(A)>P(B).$$

The third method was to choose the one with the greater or lesser difference. For example, in answering item III6 (bag A: 21 red and 8 black; bag B: 210 red and 80 black; a black marble wanted), a grade 6 student, f2063230, chose bag A and gave the following answer and reason:

> *The difference between 21 and 8 is not so big, but the difference between 210 and 80 is very big. Most of the marbles picked out from bag B should be red, not black. Therefore, I believe option (a) is correct.*

These previous examples provide an indication of the way students' intuition was used in one-stage items and the discussion will now look at two-stage items, which were more difficult.

The most widely used intuitive method in solving two-stage chance comparison items was the compound approach where students split a two-stage experiment into two one-stage experiments. They believe the combination of the most likely outcome at each stage has the greatest chance. This means, if outcome A is most likely in stage one and outcome B is most likely in stage two, then the compound outcome A and B is the most likely outcome in an experiment involving the two stages. This misconception is of particular interest in this study and, for convenience, it is referred to as the compound approach. For example, in item IV5 it was stated that bag A contains 8 red and 16 black marbles and bag B contains 50 red and 70 black marbles. The students were asked if they pulled a marble out of each bag which of the following statements was correct.

a) it is most likely that both marbles are red
b) it is most likely that both marbles are black
c) it is most likely that one marble is red and the other one is black
d) it is impossible to indicate which one is the most likely among the three outcomes

A grade 12 student, i6122470 gave the following answer and reason:

Answer: *(b)*
Reason: *In bag A, it is more likely to pull out a black marble. In bag B, it is also more likely to pull out a black marble. So it is most likely to pull out two black marbles.*

Item IV10 was a similar task but the context changed to spinning the arrowheads of two spinners. In spinner A, one sector with a central angle of 120° was coloured red and the rest part was coloured blue. In spinner B, one sector with a central angle of 300° was coloured red and the rest part was coloured blue. A grade 8 student, h4083431, gave the following answer and reason:

Answer: *(c) it is most likely that one arrowhead stops in the red part and the other arrowhead stops in the blue part*
Reason: *In spinner A, blue area is bigger than the red area. In spinner B, blue area is smaller than the red area. So I believe it is most likely that one arrowhead stops in the red part and the other arrowhead stops in the blue part.*

Although his answer is correct he used the compound approach.

The second method for two-stage items was based on a comparison of the total amount (marbles or area). This means that the students compare the total of different amounts (for example, the total number of the red marbles versus the total number of the black marbles). A grade 8 student, h5081441, used it in solving item IV10 mentioned above:

Answer: *(a) it is most likely that both the arrowheads stop in the red part*
Reason: *Because the total area of the red parts on the two spinners is bigger than that of the blue parts. Both of the other two likelihood are smaller than this one.*

However, many other students did not explain very clearly which method, the first one or the second one, they actually used. If both methods led to the same conclusion, it made coding difficult. For example, in answering item IV5 mentioned before (bag A contains 8 red and 16 black marbles and bag B contains 50 red and 70 black marbles) a grade 12 student, i6122511, gave the following answer and reason:

Answer: *(b) it is most likely that both marbles are black*
Reason: *Because there are more black marbles.*

The conclusion of that it is most likely that both marbles are black can result from using either method one or method two. Unfortunately, the student's reason is unclear. Therefore, all such responses were labelled as using either the first or the second method.

Four students applied the third method that if one amount is much more in one bag/spinner and, at the same time, it is not less in the other bag/spinner, this amount determines the compound outcome. For example, a grade 8 student, d4083221 gave the following answer and reason to item IV7. Item IV7 was parallel to IV5 but the number of marbles in two bags was changed to 8 red and 16 black versus 500 red and 100 black.

Answer: *(a) it is most likely that both marbles are red*
Reason: *This is because for bag A, as long as you correctly pull out one red marble out of two marbles, you will succeed. For bag B, the ratio between the number of red marbles and black marbles is 5:1, so it is most likely.*

Clearly, he did not use the compound approach as his conclusion was not (c), but it is unclear whether the strategy is different conceptually from the second method or just a more sophisticated variation.

The last method used by the students in this group was that the compound outcome that involves different results has more chance. This means identical results appear less often than different results. A grade 12 student, d7123160, answered item IV7 in this way:

Answer: *(c) it is most likely that one marble is red and the other one is black*
Reason: *No matter what the difference between red marbles and black marbles is, no matter the black marbles have more chance to be pulled out, but one red and one black has even more chance.*

Frequency

In this study, students' own methods of chance comparison could be observed in six one-stage items and eight two-stage items. Questionnaires E, F and G include the six one-stage items (two for each questionnaire). Questionnaires A, B, C, D, H and I include the eight two-stage items (one for A-D, two for H, two for I). It means that for students writing questionnaires E, F and G only had the opportunity to use their own methods in one-stage items. For all the other students they only had the opportunity to use it in two-stage items. The data are summarised in Tables 4.18 and 4.19.

One-stage items. An analysis of the information in Table 4.18 shows that the method involving the absolute amount was used slightly more often than the other two methods in solving one-stage items.

Table 4.18 Number of students with own methods on 6 one-stage chance comparison items

Total number of students in each stream of each grade who were given the items		Grade 6*		Grade 8*		Grade 12*	
		Adv 14	Ord 43	Adv 34	Ord 36	Adv 33	Ord 30
Number of students with *The one.....has more chance*	With the greater absolute amount	1 (7)	7 (16)	3 (9)	3 (8)	1 (3)	0 (0)
	With the greater/ lesser total	1 (7)	2 (5)	1 (3)	2 (6)	1 (3)	1 (3)
	With the greater/ lesser difference	0 (0)	2 (5)	3 (9)	1 (3)	0 (0)	0 (0)

* Figures in brackets represent the percentage of students using the misconception for the corresponding cohort.

Actually, the number of instances of the method involving the absolute amount might be greater than that shown in the table. This is because in two spinner items (items III8 and III10) the spinner with the greater absolute amount (larger area of blue) was just the spinner with the greater ratio of blue area to red area. Since the calculation of the probabilities was not required in the items the absolute amount approach in these two items could not be coded as using the misconception. Consider a grade 6 student, g2062281m, who answered items III7 and III8. Item III7 presented the student with two bags, bag A with 8 red and 16 black marbles and bag B with 500 red and 100 black marbles. He was asked to choose which bag had the greater chance for picking out a black marble. He selected bag B and explained that

> *This is because bag B has more black marbles. Compared to bag A, even though its ratio is smaller, it is superior in the number of black marbles after all.*

He applied the method involving the absolute amount. For item III8, spinner A (120° red and 240° blue) and spinner B (150° red and 210° blue) were given. The student chose spinner A as the one that gave the greater chance of the arrowhead stopping in the blue part and explained that *the spinner A has a bigger sector coloured by blue.* Probably, he used the same method, the absolute amount method in solving both problems, but his second response was not coded as using the misconception.

The conclusion that students, especially younger students, use the method involving the greater absolute amount more often is consistent with the findings of Green (1983b).

Two-stage items. The results presented in Table 4.19 show that the compound approach is a major intuitive method used by the students in all

Table 4.19 Number of students with own methods on the 8 two-stage chance comparison items

<table>
<tr><td colspan="2" rowspan="3">Total number of students in each stream of each grade who were given the items</td><td colspan="2">Grade 6*</td><td colspan="2">Grade 8*</td><td colspan="2">Grade 12*</td></tr>
<tr><td>Adv</td><td>Ord</td><td>Adv</td><td>Ord</td><td>Adv</td><td>Ord</td></tr>
<tr><td>36</td><td>81</td><td>67</td><td>72</td><td>64</td><td>57</td></tr>
<tr><td rowspan="5">Number of students with The one… … has more chance</td><td>(1) involves the more likely outcome in each stage (compound approach)</td><td>16(44)</td><td>16(20)</td><td>12(18)</td><td>12(17)</td><td>19(31)</td><td>23(40)</td></tr>
<tr><td>(2) with the more totals</td><td>1(3)</td><td>1(1)</td><td>2(3)</td><td>0(0)</td><td>1(2)</td><td>0(0)</td></tr>
<tr><td>Either (1) or (2)</td><td>6(17)</td><td>10(12)</td><td>3(5)</td><td>5(7)</td><td>6(9)</td><td>4(7)</td></tr>
<tr><td>Much more in one bag/ spinner and not less in the other bag/spinner</td><td>0(0)</td><td>1(1)</td><td>1(2)</td><td>1(1)</td><td>1(2)</td><td>0(0)</td></tr>
<tr><td>Involves different results</td><td>0(0)</td><td>0(0)</td><td>0(0)</td><td>1(1)</td><td>0 (0)</td><td>2 (4)</td></tr>
</table>

* Figures in brackets represent the percentage of students using the misconception for the corresponding cohort.

grades and streams. It appeared that grade 6 and 12 students used the compound approach more often than grade 8 students. There was virtually no difference between advanced schools and ordinary schools for grade 8. In order to explain the fact that there were less compound approach responses observed in grade 8 and in ordinary grade 6 students, students' answers to the eight tasks were classified in detail in Table 4.20, where either (1) or (2) responses were named as the unsure compound approach. It was found that for ordinary grade 6 and advanced grade 8 students, they used other misconceptions, such as chance cannot be measured and gave correct responses more often than their corresponding cohorts. Ordinary grade 8 students used the misconceptions such as subjective judgements (physical properties based) or I feel that… more often in their responses. This might explain why there were fewer compound approach responses in these students' replies.

Table 4.20 Number of students using different methods in answering the 8 two-stage chance comparison items

Total number of students in each stream of each grade who were given the item	Grade 6* Adv 36	Ord 81	Grade 8* Adv 67	Ord 72	Grade 12* Adv 64	Ord 57
Compound approach	16 (44)	16 (20)	12 (18)	12 (17)	19 (30)	23 (40)
Unsure compound approach	6 (17)	10 (12)	3 (4)	5 (7)	6 (9)	4 (7)
Chance cannot be measured	10 (28)	26 (32)	28 (43)	20 (28)	7 (11)	12 (21)
Correct response	1 (3)	11 (14)	11 (16)	2 (3)	16 (25)	9 (16)
Others	3 (8)	18 (22)	13 (19)	33 (45)	16 (25)	9 (16)

* Figures in brackets represent the percentage of students using the misconception for the corresponding cohort.

Data and Context

The effect of data and context on each method is discussed next. The data is summarised in Tables 4.21 and 4.22.

<u>One-stage items</u>. The effect of context on eliciting the first method, using the absolute amount, was obvious. The items presented in numerical form (marble items) resulted in greater use of the misconception more than those given in visual form (spinner items). However, the effect of data was not observed as four or five students used it in each marble problem. Two students used the method twice, so the total number for the first row is two more than the data shown in Table 4.18.

Table 4.21 Number of students using different methods in answering 6 one-stage chance comparison items

Number of students with *The one ……has more chance*	Marble item			Spinner item		
	III5	III6	III7	III8	III9	III10
(1) With the greater absolute amount	5	4	4	0	4	0
(2) With the greater/lesser total	1	4	0	0	3	0
(3) With the greater/lesser difference	1	5	0	0	0	0

The effect of data on eliciting the second method, the greater or lesser total amount (marbles or area) has more chance was clear, as seven out of the eight responses were observed in parallel items III6 and III9. The two items have the same data (21:8 and 210:80 vs. 270°: 90° and 270°: 90°) but different contexts. It seemed that when the composition of two bags/spinners is parallel, the method of comparing the total was more common.

It seems that both context and data played a role in students' using the third method that involves difference. Five out of the six responses were observed in item III6. Its data are given in numerical form (marble item) and the composition of the two bags looks alike (21:8 and 210:80). It seemed that such a condition is more likely to result in the strategy.

In considering this analysis it should be noted that the total number of responses was small so any conclusions here are tentative and should be considered in this light.

Two-stage items. Looking at Table 4.22, no pattern was found indicating that context played a role in students using the compound

Table 4.22 Number of students using different methods in answering the 8 two-stage chance comparison items

Number of students with *The one ……has more chance*	Marble item				Spinner item			
	IV2	IV5	IV6	IV7	IV4	IV8	IV9	IV10
(1) compound approach	4	16	13	21	2	13	15	19
(2) with the more totals	1	0	1	0	0	0	0	3
Either (1) or (2)	0	4	17	0	0	8	5	0
Much more in one bag/spinner and not less in the other bag/spinner	0	0	0	3	0	0	0	1
Involves different results	0	1	1	1	0	0	0	0

approach, but the effect of data was observed. When the compositions of two bags/spinners are equal, as in items IV2 and IV4, very few (2 or 4) compound approach responses were observed. Most incorrect responses in this situation were that it is impossible to indicate which one is the most likely among the three outcomes. When the compositions are not equal but close, as in items IV5 and IV8, IV6 and IV9, more compound approach responses were observed (ranging from 13 to 16). When the compositions were quite different from each other, as items IV7 and IV10, even more (19 or 21) compound approach responses were observed. Actually, using the compound approach in items IV6 and IV9, IV7 and IV10 can result in the correct answer, but the reasoning is wrong. Since five students used the method twice, the total number for the first row is five more than the data shown in Table 4.19.

As the other methods were observed very little in this study, the effect of data and context on these methods is not discussed.

Other

The compound approach splits two-stage experiments into two independent one-stage experiments (which can be extended to experiments with more than two stages). Suppose a and b are two possible outcomes of the first stage and c and d are two possible outcomes of the second stage. According to the compound approach if

(1) a and b are equally likely to happen and c and d are also equally likely to happen, then outcomes of two-stage task such as *a and c*, *b and c*, *b and d* and *a and d* are equally likely to happen.

(2) a and b are equally likely to happen, but c is more likely to happen than d, then *a and c* or *b and c* are more likely to happen than *a and d* or *b and d*, but *a and d* and *b and d* are still equally likely to happen.

(3) a is more likely to happen than b, c is also more likely to happen than d, then outcome *a and c* is the most likely one to happen.

Fischbein et al. (1991) and Lecoutre and Durand (1988) found the error described in (1) above when they analysed equiprobable responses. In this study, the situations were extended to other two more general situations and it was found that many students, independent of whether or not they had been taught probability, used the splitting thinking very often and consistently.

For example, student a2063011m, a grade 6 student was interviewed regarding his answer to item IV6. The item involved two bags, one containing

21 red and 8 black marbles and the other 210 red and 80 black marbles. The experiment is to pull a marble out of each bag. The student was asked to indicate which of the following statements is correct.

a) it is most likely that both marbles are red
b) it is most likely that both marbles are black
c) it is most likely that one marble is red and the other one is black
d) it is impossible to indicate which one is the most likely among the three outcomes

In his written answer he has chosen (a) indicating that it was *because there were more red*. The following is part of the interview.

Interviewer: *You chose (a), as there are more red marbles?*
Student: *Yes.*
Interviewer: *Do you want to make any changes to your written answer now?*
Student: *No.*
Interviewer: *OK. Now I ask you another question, please tell me your answer to the new question. I exchange the colour of the marbles in bag A but don't change the colours in bag B. So now there are 8 red and 21 black marbles in bag A and 210 red and 80 black marbles in bag B. Which option will you choose now?*
Student: *I prefer to say it is most likely that one [marble is] black and [the other] one [is] red.*
Interviewer: *If in bag A there are more red marbles than black marbles but in bag B there are more black marbles than red marbles, which option will you choose now?*
Student: *It is most likely that one [marble is] red and [the other] one [is] black.*
Interviewer: *How about in both the bags there are more black marbles than red marbles?*
Student: *Double black.*

Based on his written answer, it was unclear whether he used the compound approach or the totals approach. The interview showed he really based his decision on the compound approach.

There were 34 responses where it was unclear as to whether they should be labelled as using the compound approach or the totals approach. Six of these students were interviewed. Based on the interviews, it was determined that four out of the six students were actually using the compound approach, for example, the student a2063011m mentioned above. But his classmate, a2063021m, was actually using the totals approach. When the same question about exchanging the marbles' colours in bag A and leaving bag B unchanged was asked, he said

Student:	*I still choose (a) as 210+8 > 80+21.*
Interviewer:	*Could you tell me why you calculate the total of red marbles and the total of black marbles?*
Student:	*As the problem is the same as if you put all the marbles in one bag and pull out two marbles.*

Summary

This group discussed students' own methods in dealing with chance comparison items. All the three methods used in solving one-stage items referred to in the literature (for example, Green 1983b) were observed in this study.

The compound approach was one of the common misconceptions observed in this study but has rarely been studied systematically in the literature. For example, Lecoutre and Durand (1988) and Fischbein et al. (1991) did mention the idea but only focused on equiprobable answers. In this study, the situations were extended to two other more general situations. It appears that the compound approach is an important misconception when students try to solve two-stage problems.

Group 12: Taking different order as the same

The literature reports, "there is no natural understanding of the fact that, in a sample space, possible outcomes should be distinguished and counted separately if the order of their elementary components is different"

(Fischbein et al. 1991). This conclusion is supported by other research. The literature also indicates that with the help of concrete settings, such as the lining up of three children or the dressing of toy bears in all possible combinations of tops and pants, most pupils were able to notice the order (Green, 1982; English, 1993). In this study most of the items used were very simple experiments with each stage only having two possible outcomes.

Examples

This misconception could be observed directly in two parallel items (IV1 and IV3). In item IV1, the following table where the result for outcome 1 was given and the students were asked to list all the outcomes when pulling out one marble from each bag (both contain two white and two black marbles). A grade 6 student, a2062030, gave the following answer:

	Bag A	**Bag B**
Outcome 1	**White marble**	**White marble**
Outcome 2	Black marble	Black marble
Outcome 3	White marble	Black marble

She and another five students from grades 6 and 8 missed one possible outcome involving the reverse outcome for the white and black marbles. Only one grade 12 student made the same mistake. However, in more complicated items, IV2, IV7, IV9 and IV10, another six grade 12 students took the different order as the same in their calculations. For example, student c5121140, gave the following reason for her answer to item IV9 where two different sized spinners were involved. Each spinner has a sector with a central angle of 90° that was coloured blue and the remainder coloured red.

The probability of both arrowheads stopping in the red parts is 3/4×3/4=9/16. The probability of both arrowheads stopping in the blue parts is 1/4×1/4=1/16. The probability of one arrowhead stopping in the red part and one arrowhead stops in the blue part is 3/4×1/4=3/16. So (a) is the most likely one.

Her conclusion that double red is more likely is correct but the calculation is incorrect.

Frequency

Table 4.23 Number of students taking different order as the same in answering 10 items in Category IV

Item	List all		Two-stage chance comparison experiment							
	IV1	IV3	IV2	IV4	IV5	IV6	IV7	IV8	IV9	IV10
Grade 6	3	0	0	0	0	0	0	0	0	0
Grade 8	3	0	0	0	0	0	0	0	0	0
Grade 12	0	1	2	0	0	0	1	0	2	2

Table 4.23 shows that the younger students only used the misconception in marble item IV1 when they were asked to list all possible outcomes of the experiment. None of them made the error in the parallel spinner item IV3. It seems reasonable to conclude that the very small number of students using this misconception was mainly due to the simple and concrete context of the items. About 12% of the grade 12 students correctly answered the two-stage chance comparison items basing on calculating probability. As shown in previous example, all the six students (one student used it twice) in Table 4.23 calculated the probability of identical results (for example, double red) correctly but made the error in calculating the probability of different results (for example, one red and one black). Whether this was caused by the misconception or a problem of incorrectly interpreting that option (c) means only one possible outcome (pulling out a white marble from

bag A and a black marble from bag B) but not two possible outcomes is unclear. None of the six students were interviewed.

Summary

Since the number of students using the misconception is so small, a discussion of the grade, stream, role of data or context was not feasible. Also, in light of the small numbers the conclusions given in the discussion of this misconception should be considered very tentative. The data show that the error exists as has been found in the literature for Western students.

Group 13: Misuse or extend conclusions inappropriately

The outcome approach and equiprobability can be considered as strategies that students have generated based on their own experience. In both cases they apply them to probability situations as if they were principles for interpreting or describing the probability of an outcome. These two specific strategies occurred often enough to be discussed in their own categories. However, a few students misused or extended inappropriately, other approaches they have learned and these responses have been grouped together and discussed here.

Examples

Ten students in this study misused three principles that they had learned. They are the pigeonhole principle, the law of large numbers and the probability of getting one head and one tail is twice the probability of getting two heads or that of getting two tails when two fair coins are flipped.

Three younger students, without school-based instruction in probability, misused the pigeonhole principle, namely if you have 7 pigeons and put them into 6 pigeonholes, then it is certain that there is one pigeonhole containing two pigeons. The principle can be used in an experiment without replacement but these students ignored the role that the condition with or without replacement has in discussing probability. They misused it in experiments involving rolling a die or pulling out marbles from a bag with replacement. For example, a grade 8 student, d1082160, gave the following reason to explain the number 2 as a possible outcome of rolling a die:

> *There are six sides on a die. Each number has 1/6 chance to be rolled. So, if you roll 7 times, it is certain you will get a 2. But if you only roll once, you may get a 2, may get any other one of the five numbers.*

Three grade 12 students who had received instruction in probability misused the law of large numbers. Two of them believed that if an observed frequency corresponds to the theoretical probability in a short run of an experiment, then such a correspondence may not happen after a long run. So they avoided the option that matches the theoretical probability value. Another grade 12 student, f6122341, misused the law of large numbers in answering item III6. The item posed a situation where there were two bags. Bag A contained 21 red and 8 black marbles and bag B contained 210 red and 80 black marbles. The student was asked to choose a bag that gives the greater chance of pulling out a black marble.

> Answer: *(b) the likelihood of pulling out a black marble from bag A is less than that from bag B*
>
> Reason: *P(bag A)=8/29, P(bag B)=80/290=8/29. The two probabilities are equal. However, according to statistic conception, the larger sample size brings the less error.*

His classmate, student h6122411m, misused the conclusion that one head and one tail is the most likely outcome when flipping two fair coins in solving item IV10. Two spinners were given in the item, one had a red sector with a central angle of 120°, the other had a red sector with a central angle of 300°, and the remaining parts of the two spinners were coloured in blue.

Answer: *(c) it is most likely that one arrowhead stops in the red part and the other arrowhead stops in the blue part*

Reason: *Because the likelihood that one arrowhead stops in the red part and the other arrowhead in the blue part is twice the likelihood of both arrowheads stop in the red parts or that of both in the blue parts*.

Frequency

From examining the information in Table 4.24, it seems that the main problem for the younger students was being unaware the role of the concept of replacement plays in dealing with probabilistic problems. For the older students who had some knowledge of probability, they might need more experimental experiences to help them understand the theory better.

Table 4.24 Number of students who misused or extended conclusions inappropriately

Total number of students in each stream of each grade	Grade 6		Grade 8		Grade 12	
	Adv 50	Ord 124	Adv 101	Ord 108	Adv 97	Ord 87
Pigeonhole principle	1	1	1	0	0	0
Law of large numbers	0	0	0	0	2	1
P(H,T or T,H) =2P(H,H) =2P(T,T)	0	1	0	0	3	0

Summary

As this group involved only ten students, to draw any major conclusions would not be valid. However, the discussion above raises some interesting questions concerning misusing or extending principles that are generated or specifically taught in other situations. It seems worth considering this area as a potential subject for further research.

Group 14: Used own methods in chance calculation

Although all the students in grades 6 and 8 had never been taught how to calculate probability, some students used calculations in their answers. Also, some grade 12 students used their own calculation methods. The common idea connecting these methods was the underlying use of ratio to measure chance.

The first part of this discussion relates to three kinds of methods used in solving one-stage items. Instead of using part-whole ratio, they tried to use (1) part-part ratio; (2) part-rough whole ratio or (3) the reciprocal as the measure of probability. The first method has been reported very often in literature and was reviewed in Chapter 2.

The next part of the discussion relates to two kinds of methods used in multi-stage items. Instead of using the additive rule or/and multiplicative rule to the probabilities obtained in each stage, they tried to obtain the probabilities of compound events by using (1) a computation involving the number of stages or (2) average. These methods were rarely reported in literature.

Examples

<u>*One-stage items*</u>. The first and the main own method is using part-part ratio. Consider item III5 which posed a situation where bag A had 8 red and 16 black marbles and bag B had 50 red and 70 black marbles. Students were asked to indicate which bag gave the greater chance for pulling out a black marble. A grade 12 student, e5121241, calculated as follows:

> *Bag A has 50% chance. Bag B has 5/7 chance. 5/7>1/2. So bag B gives more chance.*

Although he had been taught the classical probability formula instead of calculating part-whole ratio he used the part-part ratio to calculate the probabilities. It seemed that a part-part ratio method was a component of some students' schema rather than the part-whole ratio that had been taught. Another three students used both the part-whole ratio and part-part ratio in answering one item. For example, another grade 12 student, d7124131, gave the following calculation to item IV7 where bag A contained 8 red and 16 black marbles and bag B contained 500 red and 100 black marbles. He wrote that *in bag A, the probability of pulling out a black marble is 2/3. In bag B, the probability of pulling out a red marble is 1/5*.

The second method is part-rough whole ratio. This means that students used the rough whole number instead of the exact whole number in calculating probability. Item III4(3) involved a school with 200 girls and 1000 boys. Students were asked that after 35 girls' and 35 boys' names had been picked out, is it more likely or equally likely to pick out a boy's name or a girl's name next. A grade 8 student, i5081531, calculated chance in this way:

> *When you take 35 girls from the total 200, then the number is 165. When you take 35 boys from the total 1000, then the number is 965. …If calculated by percentages, that is, the probability of picking out a boy's name is 96.5%, while, that for a girl's name is only 16.5%. So boys have more chance.*

He preferred to measure chance by percentage, but thought that he only needed to change the two numbers 965 and 165 into 96.5% and 16.5%. He gave a similar answers to item III4(4) (another slip item), but no calculations in his other answers. Actually he should use the whole students number 1130 as the denominator, but possibly he used the rough whole students number 1000 instead. Based on his responses to other items it was clear that the student understood the concept of percentage. In light of this, while an explanation that he used a part-rough whole ratio calculation is tentative, it is worth considering.

The last own methods in chance calculation for one-stage items is using the reciprocal as the measure of chance. This means that if there are 8 black marbles in a bag then the chance of pulling out a black from the bag is 1/8, no matter how many other coloured marbles are put in the bag. Item III6 involved bag A containing 21 red and 8 black marbles and bag B containing 210 red and 80 black marbles. A grade 12 student, f7123290, answered that there was a greater chance of getting a black marble from bag A as *the chance of pulling out a black marble from bag A is 1/8, from bag B is 1/80.*

Fewer examples of students using their own methods in chance calculation were observed in multi-stage chance comparison items.

<u>*Multi-stage items*</u>. The first method used in multi-stage items is that after working out the probability of each outcome in each stage, students measured the probability of a compound event by a computation involving the number of stages. For example, in answering whether the outcome of none of the three numbers rolled is 6 is impossible, possible, or certain to happen, when three normal dice were rolled once, a grade 8 student, d5081190m said it was possible, because

> *The probability for one die is 1/6. So for three dice, the probability is 1/6+1/6+1/6=1/2.*

She did not realise that the chance of three 6s is much smaller than that of one 6.

For the same item, another grade 8 student, d1082160, gave another reason, namely *…as three dice are rolled, the chance of all the three numbers are 6 is 1/18. …* She divided 1/6 by 3.

The second method is taking the average value as the compound event's chance value. In item IV10, two spinners were given, one had a red sector with a central angle of 120°, the other had a red sector with a central angle of 300°, and the remaining part of the two spinners were coloured blue. A grade 12 student, h6122450, calculated the probabilities of two arrowheads stopping in the same colour parts or different colour parts as follows:

> *Red:* $\frac{120+300}{720}=\frac{420}{720}$ *One red one blue:* $\frac{240+300}{720}=\frac{640}{720}$ *Blue:* $\frac{240+60}{720}=\frac{300}{720}$
> *Therefore, I choose option "c" [it is most likely that one arrowhead stops in the red part and the other arrowhead stops in the blue part].*

Frequency

Table 4.25 shows that the most common method was using part-part ratio to measure chance. It was the only method observed in each grade and each stream with all the other methods being used less often. Two-thirds of the responses reported in Group 14 were found in solving one-stage chance comparison items.

Table 4.25 Number of students who used own methods in chance calculation

<table>
<tr><td colspan="2" rowspan="3">Total number of students in each stream of each grade</td><td colspan="2">Grade 6</td><td colspan="2">Grade 8</td><td colspan="2">Grade 12</td></tr>
<tr><td>Adv</td><td>Ord</td><td>Adv</td><td>Ord</td><td>Adv</td><td>Ord</td></tr>
<tr><td>50</td><td>124</td><td>101</td><td>108</td><td>97</td><td>87</td></tr>
<tr><td rowspan="3">One-stage items</td><td>Part-part</td><td>1</td><td>3</td><td>1</td><td>0</td><td>1</td><td>1</td></tr>
<tr><td>Part-rough whole</td><td>0</td><td>0</td><td>2</td><td>0</td><td>1</td><td>0</td></tr>
<tr><td>Reciprocal</td><td>1</td><td>0</td><td>0</td><td>0</td><td>1</td><td>2</td></tr>
<tr><td rowspan="2">Multi-stage items</td><td>Multiplied or divided by stage number</td><td>0</td><td>1</td><td>2</td><td>1</td><td>0</td><td>0</td></tr>
<tr><td>Average</td><td>0</td><td>0</td><td>2</td><td>0</td><td>1</td><td>0</td></tr>
</table>

In this study, the number of students using their own calculation methods is small. The students were not required to calculate probability and many items can be answered without doing any calculation. Also, the data presented in previous sections show that students used many other approaches rather than calculating probabilities in their answers. For example, using chance cannot be measured mathematically (Group 4), equiprobability (Group 5), positive and negative recency (Group 10) and own methods in chance comparison (Group 11). Inventing their own calculation methods is only one of many possible approaches that students could have used. Furthermore, as mentioned earlier, grade 6 and 8 students had not been taught how to calculate probability so the other methods were more likely to be used. Consequently, it is possible that incorrect calculation methods are more prevalent in the students' set of approaches than implied by the data, a situation that could be investigated in another study where students are required to include calculations.

Summary

There are 21 students who used their own calculation methods in chance comparison tasks. All the methods have an underlying use of ratio as a common theme. Using part-part ratio instead of using part-whole ratio was the most common method used by each grade and each stream. The other methods reported here were only used very occasionally and have rarely been mentioned in the literature. The use of students' own calculation methods is one possible area that might be worthy of further research.

Summary of the Chapter

Fourteen groups of misconceptions of probability are described in this chapter. Since approximately two-thirds of the students did not have any formal probability background, the most common misconceptions observed were naïve knowledge of uncertainty, simple measurement of likelihood, superficial interpretation of probability and intuitive methods of chance comparison. Table 4.26 lists the five most common misconceptions observed in this study.

Table 4.26 The top five misconceptions observed in this study

Top 5 misconceptions	Rough percentage of students clearly used it at least once
Group 6: Outcome approach	38%
Group 4: Chance cannot be measured	33%
Group 11*: Compound approach	26%
Group 5: Equiprobability	19%
Group 8: Data match or word match	13%

* Only clearly compound approach responses were counted, other own methods in chance comparison responses reported in Group 11 were not included. This note is also for the later tables.

Clearly, the first three misconceptions were much more common than the next two. The rough percentage for the 6^{th} common misconception (increasing repetition is not better for predicting) is 11% and that for the 7^{th} (example-based interpretations for possible and impossible) is 10%.

The data in Table 4.26, 4.27 and 4.28 are titled <u>rough</u> because they are based on the frequency information reported in each group of this chapter, and as mentioned earlier, not all the misconceptions' responses were counted in the previous frequency tables. For the top three misconceptions, some students (from 4 to 15) were not counted in their frequency tables as the data are based on the occurrence of the misconceptions in some not all categories. Furthermore, the frequency tables only include students who clearly used a misconception. The unsure compound approach responses in Table 4.20 were not counted in Table 4.26, 4.27 and 4.28.

Table 4.27 Rough percentage of students who used the top three main misconceptions at least once in each grade

Grade	First main misconception	Second main misconception	Third main misconception
6	Group 6: (43%) Outcome approach	Group 4: (46%) Chance cannot be measured	Group 11: (27%) Compound approach
8	Group 6: (39%) Outcome approach	Group 4: (35%) Chance cannot be measured	Group 5: (19%) Equiprobability
12	Group 11: (34%) Compound approach	Group 6: (31%) Outcome approach	Group 4: (19%) Chance cannot be measured

Examining the frequency of misconceptions that occurred at different ages and in different school streams, it was found that the two most common misconceptions, chance cannot be measured mathematically and the outcome approach, were major misconceptions for all grades and streams.

Table 4.28 Rough percentage of students who used the top three main misconceptions at least once in each stream

Stream	First main misconception	Second main misconception	Third main misconception
Adv	Group 6: (37%) Outcome approach	Group 11: (28%) Compound approach	Group 4: (29%) Chance cannot be measured
Ord	Group 6: (38%) Outcome approach	Group 4: (36%) Chance cannot be measured	Group 11: (24%) Compound approach

However, it should be noted that such a listing is not absolutely fair for each misconception observed in this study. This is because the number of items for eliciting the different misconceptions is different. Thirty items in the questionnaires were designed primarily to investigate chance cannot be measured or equiprobability. Another twenty-seven items were designed to investigate the outcome approach, but only 8 items were designed to investigate the compound approach and only 6 items for the misconception of increasing repetition is not better for predicting. Therefore, the lists might be changed if each misconception had an equal chance to be observed. In spite of this limitation, the researcher feels that the summary presented is useful in providing an overview of the situation.

The analysis of the students' responses shows that context and data play a role in eliciting some misconceptions.

Influenced by context	Influenced by data
Physical properties based judgement Equiprobability Outcome approach One trial is unrelated to other trials Interpret chance by data match or word match	Equiprobability Positive and negative recency Compound approach

This is a brief overview of the chapter. A more extensive analysis of these results as they relate to implications for teaching, and so on is included in chapter 7. The next chapter will consider the developmental levels of the students.

Chapter 5 Developmental Structure in Understanding of Probability

Chapter 5 answers the second research question, "What is the developmental structure of students' understanding of probability?" A cognitive framework describing students' hierarchical responses was generated. Since the main purpose was to investigate the developmental structure, per se, but not to assign a cognitive label to each student, the SOLO (Structure of the Observed Learning Outcome) taxonomy (Biggs & Collis, 1982; Biggs & Collis, 1991 and Collis & Biggs, 1991) was chosen as the framework. The students' responses to each item were labelled at five levels: prestructural, unistructural, multistructural, relational and extended abstract. The following description is a paraphrase of Biggs and Collis (1982) and Collis, Romberg, and Jurdak (1986), but the interpretative examples of theoretical probability were added by this researcher.

Level 0 - Prestructural (P). A response at this level indicates a refusal or inability to become engaged in the problem. It could be blank, fully irrelevant, illogical, egotistic answers or saying, "haven't done those yet".

Level 1 - Unistructural (U). A response at this level uses one, but only one, relevant operation. For example, students indicate that it is possible for a target event to happen when they find an example as evidence, but whether there is any example of other non-target events happening was not considered.

Level 2 - Multistructural (M). A response at this level uses several disjoint relevant operations, usually in sequence. For example, after students list all possible outcomes, both target and non-target, of an experiment they conclude that all outcomes have an equal chance of happening.

Level 3 - Relational (R). A response at this level contains the elements of abstract thinking. Elements are related and an integrated understanding of the information is achieved. For example, students group all the possible outcomes in favour of a target event together and use ratio as a measure of probability.

Level 4 - Extended abstract (E). The response is purely abstract thought. An abstract general principle derived from or suggested by the given information is used. For example, students successfully use the classical probability formula appropriately even though the setting is an extension of the students own real-life experiences.

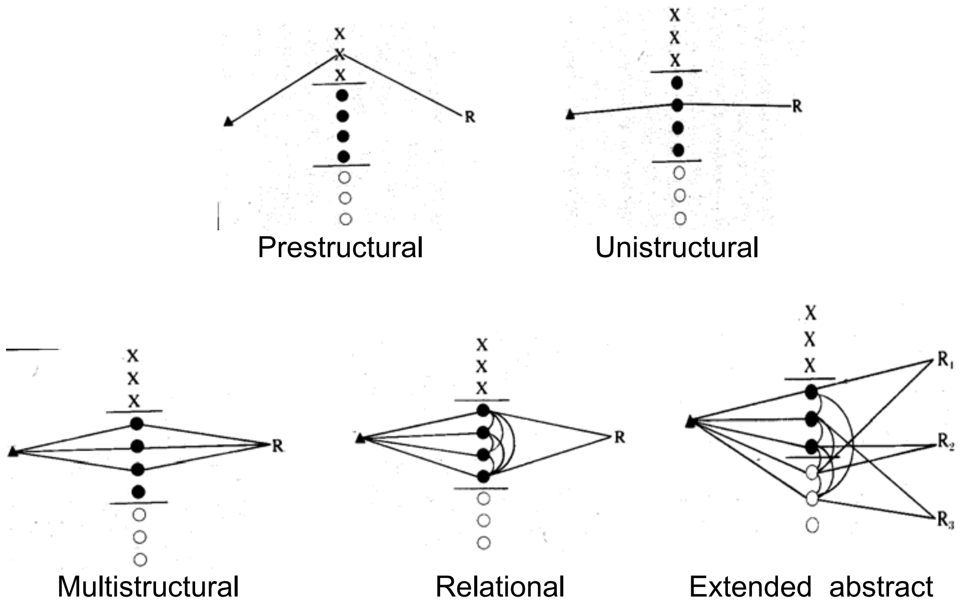
Figure 5.1 gives a diagrammatic representation of response structure.

The report on the research question includes the following parts:

Response structure in each category items. In this study, items for the questionnaires were developed under four categories. For each category responses at each level are illustrated and sample students responses are given.

Understanding indices. For each student, he or she had two understanding indices. One was a descriptive label, formed by a set of SOLO codes for all his or her written responses. The other was a numerical label, assigned by a Rasch analysis basing on the descriptive index. Questions such as whether the students' understanding improved with age or whether advanced school students have a better understanding than ordinary school students are investigated using the indices.

Figure 5.1 Response structure described by SOLO



X = irrelevant or inappropriate; • = related and given in display;
◦ = related and hypothetical, not given.

(Biggs & Collis, 1982, p. 24-25)

Other. In this part additional information from interviews is reported and discussed.

Before presenting the specific results three points should be noted. First, each student answered only one questionnaire with 9 or 10 questions, so the analysis is based on their responses only to the items that they completed. Second, as explained in Chapter 3, if a choice was selected but no reason was given, the choice was assigned the lowest level that could lead to the choice. When a student answered a different question, due to misreading or misunderstanding the question, if it was possible to code the response it was coded based on the new question. Otherwise, it was coded as missing data. The discussions on the descriptive index were based on the responses that could be coded by the SOLO model. Third, students' actual understanding levels might be underestimated as their written responses might not indicate the complexity of their actual thought process.

Response structure in each category

In this study, data on students' understanding of probability were collected when they answered items in the following categories: (1) identification of impossible, possible and certain events; (2) interpretation of chance values; (3) chance comparison in one-stage experiments; and (4) chance comparison in two-stage experiments.

Category I: Identification of impossible, possible and certain events

In this category, students were asked to indicate whether an event was impossible, possible or certain to occur in a trial. The situations were rolling dice (die) or drawing out playing cards (card). Unless indicated otherwise all responses used to illustrate the levels are responses for item I3(1).

Item I3(1) Three six-sided normal dice are rolled once. Please indicate whether the outcome "all three numbers rolled are even numbers" is impossible, possible or certain to happen. Tick where appropriate:

impossible	possible	certain
☐	☐	☐

<u>Prestructural responses</u>

According to the SOLO taxonomy, prestructural responses indicate a refusal or inability to become engaged in the problem. So, for example, blank, fully irrelevant, illogical, egocentric answers and answers showing that a student is unable to understand the meaning of impossible, possible and certain are indications of P responses. A grade 6 student, g3064340, gave the following deterministic conclusion based on her feeling, which was irrelevant and inappropriate.

> Answer: *Impossible*
> Reason: *It's impossible to make such a coincidence. So I feel it is impossible.*

Another student, g3064311m, a classmate of the above student, gave the following answer.

> Answer: *Chose both impossible and possible*
> Reason: *This is because the outcome is not decided by you, it's a possible outcome but also an impossible outcome. Further more, the die is six-sided, outcome is various.*

<u>*Unistructural responses*</u>

The SOLO taxonomy indicates that a typical response at this level uses one, but only one, relevant operation. In Piaget theory, operations mean cognitive actions, which are organised closely together into a strong structure (Piaget & Inhelder, 1975, p 248). In dealing with the identification items, students who give U responses are more likely use one relevant option, that is, they try to find example(s) of the target event occurring or not occurring. Then they achieve closure: possible/impossible because they have found evidence that the event was possible/impossible to occur. For example, a grade 6 student, g1061391, gave the following answer and reason,

> Answer: *Possible*
> Reason: *As it is possible the three numbers rolled are 2, 4, 6, all the three are even numbers.*

The student achieved closure when he found a supportive example of the target event and fortunately he got the correct answer. However, the following student h4083480, a grade 8 student, gave a wrong answer to question I3(2) because of her quick closure. The item in the question was about the event that all the numbers rolled are smaller than 7 when three normal dice were rolled once.

> Answer: *Possible.*
> Reason: *If all the three numbers rolled are 1.*

<u>*Multistructural responses*</u>

Multistructural responses use several disjoint relevant operations, usually in sequence. In answering the identification items, students who gave M responses usually listed all possible outcomes of an experiment and indicated that the target event only included part of the possible outcomes. Alternatively, they described the likelihood of the target event by using words such as higher chance, lower chance, equal chance or gave a subjective estimation for the likelihood. The preliminary measure was based on their experience or intuition. For example, a grade 8 student, g5081410, chose *possible* and gave the following reason

> *As there are three odd numbers and three even numbers on each die, that is, the probability of rolling out an even number is 50%. Then, the probability of rolling out three even numbers on three normal six-sided dice is also 50%.*

She stated the target and non-target events for each die first, then appropriately assigned equiprobability to the two outcomes. But her extension to three dice was inappropriate, based on her own method of chance calculation (taking the average value in each stage as the compound event's chance value).

Relational responses

According to the SOLO model, a relational as opposed to a multistructural response contains the elements of abstract thinking. Students who made relational responses were much more likely to be able to take all the possible outcomes that belong to target event as a group and use ratio to measure probability. For example, in answering item I1(6) where the target event was the number rolled is not 6, a grade 12 student, f5121300, chose *possible* and gave the reason that *it's possible that the number rolled is 1,2,3,4,5. The probability is 5/6.* She connected all the possible outcomes belonging to the target event together and calculated by using the classical probability formula. Other ratio methods of calculating probability of two- or three-stage events are also coded as R responses.

Extended abstract responses

An extended abstract response is purely abstract thought. At this level, the classical probability formula is used successfully when the setting is complicated. Students who gave E responses liked to work out probability by constructing an equally likely sample space or introducing abstract principles, such as the addition rule for mutually exclusive events and the multiplication rule for independent events in chance calculation. For example, a grade 12 student, g6122391, chose *possible* and gave the following reason,

> *The likelihood is $\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = 1/8$. Even though it is small, not impossible.*

Category II: Interpretation of chance values

In this category, students were asked to interpret the meaning of a specific chance value or indicate whether a prediction is accurate or not when a specific result is given. In illustrating the different SOLO levels of the students' responses, unless indicated otherwise, the following items, II2(6) and II3(2), are used:

Item II2(6) A mathematician filled a bag with black and white marbles. He didn't really know how many black marbles and white marbles were in the bag. After mixing them thoroughly, he took a look and predicted that " if I pull out a marble from the bag without looking, the chance that it will happen to be white is 30%." He pulls out a marble. The marble is white. Do you think the mathematician's prediction is accurate or not?

Item II3(2) A weather forecaster said that "tomorrow there is a 50% chance that it will rain." Which of the following has the closest meaning to "tomorrow there is a 50% chance that it will rain"?

a) it may or may not rain tomorrow. The forecaster doesn't really know what the result will be
b) suppose that there were 10 days in a year the forecaster said that "tomorrow there is a 50% chance that it will rain", and on around 5 out of the 10 days it rains the next day
c) suppose that there were 100 days in a year the forecaster said that "tomorrow there is a 50% chance that it will rain", and on exactly 50 out of the 100 days it rains the next day
d) suppose that there were 100 days in a year the forecaster said that "tomorrow there is a 50% chance that it will rain", and on around 50 out of the 100 days it rains the next day

Prestructural responses

As mentioned in the discussion of category I items, prestructural responses mainly include blank, fully irrelevant, illogical or egocentric answers. For example, a grade 6 student, g3064200, gave the following totally irrelevant reply for item II3(2),

Answer: *(b)*
Reason: *I think the second method is correct. As it is possible that around 5 days it rains once, but impossible that around 50 days it rains once.*

It seemed that she had problems in comprehending the item. She appeared to think that option (b) meant it rained around every five days and the option

(d) meant it rained around every fifty days. The comprehension problem made her unable to become meaningfully engaged in the problem.

<u>Unistructural responses</u>

Responses at this level use one and only one relevant operation. Their only relevant operation was interpreting a chance value between 0 and 1 as the target outcome may or may not happen. Since the target outcome refers to a trial, they never appear to think that the information from other repetitive trials has any relationship to the target trial, so the given information on frequency was totally ignored. They usually believed that chance cannot be measured, they did not refer to 50% or P() when they explained the meaning of P(A). It is quite likely that they had no idea that P(A)+P(A) = 1.

For example, a grade 6 student, d1061260, in answering item II3(2) chose (a) and she said that *50% only refers to the likelihood of rain on one day. It doesn't refer to the days in a year. So I choose option (a).*

The next example shows that students who give U level responses often only notice the target outcome and are unaware of any inconsistency in their replies. Item II2(3) asked students to select from the following choices which had the closest meaning of a 30% chance of pulling out a while marble.

a) the marble pulled out will certainly be white
b) the marble pulled out will certainly be black
c) suppose that the game is repeated 10 times, white marbles are pulled out around 3 times
d) suppose that the game is repeated 10 times, white marbles are pulled out exactly 3 times

A grade 12 student, b5121071m, gave the following answer and reason:

Answer: *(a)*
Reason: *As what he said is white marble not black marble.*

He thought when the target outcome (pulling out a white marble) occurs the prediction is accurate. He did not appear to realise the equivalence between the two predictions of having a 30% chance of pulling out a white marble and having a 70% chance of pulling out a black marble. Apparently he was unaware that if the given prediction was stated in the alternative way and he consistently used the same reason, he would reject the present option and choose option (b) instead for the same item, since the mathematician would have referred to a black marble not a white marble.

Multistructural responses

Responses at this level show that the students use several disjoint relevant operations, usually in sequence. At this level, students either believed or did not believe that one trial has any relationship to other trials. For students who believed that one trial has no relation to other trials, they usually used the outcome approach, that is, based their judgement on the comparison of P(A) and 50% or P(A) and P($\bar{A}$). Alternatively, they descriptively interpreted the chance value between 0 and 1 as the target outcome was likely/unlikely to happen. For the students who believed that the chance in a specific trial can be explained by the observed frequency, they usually accepted the fact that a more likely outcome does not occur in one trial, but they did not suggest any plan to examine the accuracy of the given prediction. When they chose the option(s) involving repetitions their selection was simply based on information matching or their own very limited experience with sampling.

For example, a grade 8 student, h4084470, gave the following answer and reason for item II2(6) using the outcome approach,

Answer: *Inaccurate*
Reason: *If the chance as he said is 30%, it is lower than a half, means the chance for pulling out a white marble is very small. However, the fact shows [the marble] is a white marble. So his prediction is inaccurate.*

She believed that if an outcome had been predicted with only a small chance (lower than 50%) of happening but actually it happened, it meant the prediction was inaccurate.

Another student, a grade 6 student, h2063330, did not use the outcome approach. She gave the following answer and reason for item II2(6),

Answer: *Accurate*
Reason: *This is because even a 1% chance is possible to happen, let alone 30%?*

She knew that was possible, even though unlikely, to pull out a white marble but she did not mention any further plan to check the prediction, which was a distinction between M responses and R responses.

The example of making a choice based on information matching without understanding of the rationale for repetition was given by student d1082170, a grade 8 student. She answered item II3(2) in this way:

Answer: *(b) and (d)*
Reason: *This is because the meaning of the sentence is that the possibility of rain and no rain is half-and-half, both options (b) and (d) mean that.*

She chose the two options because both of them match the 50% chance. She did not think increasing repetition was better for predicting.

Relational responses

Responses at this level show that the students' thinking contains the elements of abstract thinking such as the more repetitions the more reliable the estimation. For example, a grade 12 student, d6122180, gave the following answer and reason for item II3(2),

Answer: *(d)*
Reason: *50/100 $\times$100% = 50%, select the option with the most days for the sake of accuracy.*

Although both (b) and (d) match the 50% chance, she indicated that she preferred the option with the bigger number of repetitions as the larger sample is a more reliable predictor than the smaller one.

In answering open-ended items, students who gave R responses were more likely to suggest checking the marbles numbers or asking the mathematician to make a few more trials to verify the accuracy of his prediction. For example, a grade 12 student, h5121371, gave the following answer and reason for item II2(6),

Answer: *May be accurate*
Reason: *If the ratio of the number of the white marbles to the number of the black marbles is 3:7, the prediction is accurate.*

Another grade 12 student, h6122411m, suggested having more trials,

Answer: *Accurate*
Reason: *I believe his prediction is accurate, because if he did not have full confidence, he would not predict. However, the result was by chance, so couldn't be taken as strong evidence. He must pull out more marbles to support him.*

In the interview when he was asked what was his opinion if the chance was not 30% but 80%, he said he still believed the prediction was accurate. He answered accurate and not inaccurate because he believed that the mathematician could make a quick estimation.

Both the responses were coded as R not E. This was because for the first student he was able to put all white marbles as a group and use part-whole ratio to verify the chance of pulling out a white marble. Although the method of checking the marbles numbers was not indicated in the item, the condition that the mathematician did not really know how many black marbles and white marbles were in the bag was stated so it prompted students to find the method to confirm their conclusion. For the second student, in order to verify the mathematician's prediction, he only asked for *more* trials not *a large number of* trials, which was the distinction between R responses and E responses.

<u>*Extended abstract responses*</u>

Students who gave E responses were more likely to suggest making a large number of repetitions to examine the accuracy of a prediction. The abstract general principle, the law of large numbers, was applied automatically. For example, a grade 12 student, h5121411, in an advanced school, gave a good answer for item II2(6).

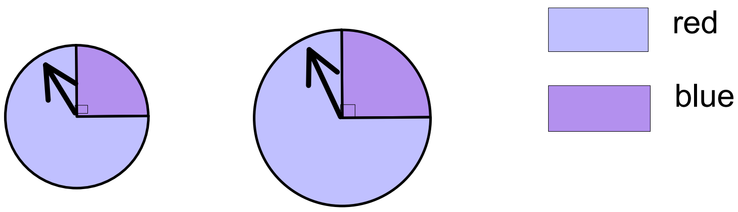
> Answer: *Accurate or not is not sure yet*
> Reason: *As a lot of experiments are needed to make a judgement whether his prediction is accurate or not. Or if the number of white marbles and black marbles is known, we can make the judgement.*

Category III and IV:

Both category III and IV contain chance comparison items. Since there are many commonalties in the responses for items in the two categories, especially at the lower levels, they are discussed together.

Category III items were one-stage tasks such as drawing one slip from a box or drawing a single marble from a choice of bags. Students were asked to compare the likelihood of drawing a boy's or a girl's name or drawing a marble of a given colour from one of the two bags. Category IV items focused on two-stage experiments such as spinning the arrowheads of two spinners or drawing one marble from each of two bags. In discussing the students' responses at each level the following items III9 and IV7 are used.

Item III9



spinner A spinner B

Spin each spinner's arrowhead with all your strength. Suppose you want the arrowhead to stop in the blue part. Which statement below is correct?

a) the arrowhead of spinner A is more likely to stop in the blue part than that of spinner B
b) the arrowhead of spinner A is less likely to stop in the blue part than that of spinner B
c) the arrowhead of spinner A is equally likely to stop in the blue part as that of spinner B
d) it is impossible to compare the likelihood of the two outcomes

Item IV7 There are 8 red marbles and 16 black marbles in bag A. There are 500 red marbles and 100 black marbles in bag B. Mix the marbles in each bag thoroughly. Put your hands in two bags and pull out a marble from each bag without looking. Which statement below is correct?

a) it is most likely that both marbles are red
b) it is most likely that both marbles are black
c) it is most likely that one marble is red and the other one is black
d) it is impossible to indicate which one is the most likely among the three outcomes

Prestructural responses

As with the other categories, blank, fully irrelevant, illogical, egotistic answers or something like, "I haven't done those yet" are labelled as prestructural responses. For example, a grade 12 student, e7124181, gave the following irrelevant reply to III9. He ignored the given condition of spinning with all your strength and tried to use strength to control the result.

> Answer: *(a)*
> Reason: *Spinner A has a smaller arrowhead, relatively its inertia is smaller. So it's easier for strength control.*

As two-stage items are the most sophisticated items used in this study, some students felt that their intuitions could not help them deal with two-stage tasks. For example, a grade 8 student, d1082191 gave the following answer and reason to IV7,

> Answer: *(d)*
> Reason: *It's impossible for me to indicate which of the three is the most likely one. I might lose my sixth sense.*

<u>*Unistructural responses*</u>

Responses at this level show that the students use only one relevant operation. For those students who thought chance cannot be measured mathematically and then cannot be compared, their only relevant operation was to explain chance as something can happen at any time. For example, a grade 6 student, e3064240, gave the following answer and reason for item III9.

> Answer: *(d)*
> Reason: *Spinner A is smaller than spinner B. It doesn't mean it is less likely to stop in the blue part. Spinner B is bigger than spinner A. It doesn't mean it is more likely to stop in the blue part, either.*

She believed the final positions of the arrowheads were unpredictable so chance comparison was impossible.

For the students who made chance comparison, their only relevant operation was to compare the spinners' physical properties like the following grade 8 student, e5081260, replied for item III9,

Answer: *(c)*
Reason: *The two angles between the arrowheads and the blue parts on the two spinners are almost the same*.

<u>*Multistructural responses*</u>

Responses at this level show that the students use several disjoint relevant operations, usually in sequence, but the pieces of information are not connected. While students who gave U responses thought it was impossible to compare chance, students who gave M responses were more likely to assign an equal chance to every possible outcome, no matter how many possible outcomes form the target event. Alternatively, they used their own non-proportional methods in chance comparison. When solving two-stage experiments they were more likely to ignore the order of the stages.

For example, in answering item III9 a grade 8 student, e1082230, assigned a 50% chance to each coloured part even though ¾ of the area on each spinner was red,

Answer: *(c)*
Reason: *No matter red or blue, each has 50% chance.*

She knew that both stopping in the red or blue parts were possible outcomes and then assigned an equal chance to each outcome.

Other students who gave M responses, based the second operation (chance measurement) on their real-life experience, namely, that the greater the amount, the greater chance. Their invented non-proportional methods for chance comparison that are reported in chapter 4 (Group 11). For example, a grade 6 student, e2063161m, selected (b) as his answer to item III9 and gave the reason that *spinner B has more blue part, so I chose (b).* He thought the spinner with the larger area of blue has the greater chance to make its arrowhead stop in the blue part.

For item IV7 a grade 8 student, d5081230, gave the following compound approach reply,

> Answer: *(c)*
> Reason: *In bag A, black marbles are more than red marbles, so it is more likely to pull out a black marble. In bag B, red marbles are more than black marbles, so it is more likely to pull out a red marble. Combine them, so it is most likely that one marble is red and the other one is black.*

She believed that the compound outcome, which is the combination of the more likely outcome at each stage, had a greater chance. Since the above two responses did not connect the data in each spinner/bag by using the concept of ratio, they were coded as M responses.

Responses that took a different order as the same were mainly observed in the two outcome listing tasks. Consider item IV1, where the following table with the first row was given and the students were asked to go on listing all the outcomes when pulling out one marble from each bag that both contains two white two black marbles. A grade 6 student, a2062030, gave the following answer:

	Bag A	**Bag B**
Outcome 1	**White marble**	**White marble**
Outcome 2	*Black marble*	*Black marble*
Outcome 3	*White marble*	*Black marble*

She missed one possible outcome, a black marble from bag A and a white marble from bag B. Probably, she took (white, black) and (black, white) as the same outcome of the two-stage experiment, because the role of order had not been internalised.

Relational responses

Responses at this level show that the students' thinking contains the elements of abstract thinking. The transition into relational level is marked by students using proportional reasoning. Students find ratio is an indicator of chance comparison. However, instead of calculating part-whole ratio some of them used part-part ratio to measure probability or, even though they calculated the part-whole ratio they do not indicate the ratio is the probability (perhaps they do not know; this will be discussed in next chapter). For example, a grade 6 student, e2063181, gave the following answer and reason for item III9,

> Answer: *(c)*
> Reason: *The two spinners have a same ratio as both of them have a sector with a central angle of 90°, although they have different radii. We shouldn't say spinner B with a larger blue area so its chance is greater than spinner A.*

It seems that the conclusion of equiprobable was based on the conclusion that they have the same ratio, even though the value of the common ratio was not explicitly stated.

The following grade 8 student, d1082160, although she still split a two-stage task into two independent one-stage tasks, based her answer to the one-stage tasks on proportional reasoning.

> Answer: *(c)*
> Reason: *The marbles in bag A could be thought of as 2 black marbles with one white [red] marble. In bag B [they] could be thought as 5 red marbles with 1 black marble. So a black marble is more likely in bag A and a red marble is more likely in bag B.*

Extended abstract responses

At this level, students making extended abstract responses were much more likely to be able to make the comparisons after working out the probabilities of each outcome. For example, a grade 12 student, d5121191, gave the following answer and reason for item IV7,

Answer: *(c)*
Reason: *This is because,*
Bag A: Red 8/24, Black 16/24 16/24>8/24
Bag B: Red 5/6=20/24, Black 1/6=4/24 20/24>4/20

It seemed that he still split this two-stage task into two independent one-stage tasks, but his answer shows that his understanding of the classical definition of probability was at the E level.

Understanding level indices

Each student has two indices, one is descriptive and one is numerical. The descriptive index is a set of SOLO levels attached to all his or her written responses. It is an initial index with the numerical index being derived from it. Before discussing the details a cautionary note on ceiling level needs to be included.

Table 5.1 Numbers of items in each category that have a ceiling level of M, R and E

Ceiling level	M	R	E	SUM
Category 1	14	2	8	24
Category 2		18	9	27
Category 3		16	6	22
Category 4		2	8	10
Total				83

All the 83 items used in the main study were in one of the four categories. In category I, 14 items have an M ceiling level. This means that responses to these 14 items can only be at the P, U or M level. For example, item I1(1), where the students were asked whether it is impossible, possible or certain that when an ordinary die is rolled once the number rolled is an even number, the nature of the task means that R and E level responses are not possible. When a student answered possible with the reason that the probability for the target event is ½, the response was coded as an M level response. The ceiling level on some items means that when comparing data from items or categories, there is an operational ceiling on the number of responses that could be at an extended abstract level, say. However, this operates in **a comparable manner across all streams and grades**, so it does not impact the general conclusions regarding the differences between streams or grades.

The descriptive index

In order to investigate students' understanding in each category the students' SOLO level responses were grouped by category. The following four tables show the number of hierarchical responses observed in each category by grade and stream, but those responses that were impossible to code are not included. For example, in Table 5.2, 134 of the responses from advanced grade 6 students to category I items were coded by SOLO level. Seven of the 134 (5%) were P responses. Since as shown in Table 5.1 the number of items in each category ranged from 10 to 27, the total number of responses in each table varies considerably.

In answering category I items, the data from Table 5.2 show that the percentage of students at the P level stays fairly constant, except that it is considerably higher for ordinary grade 6 students. Most responses for all grades and streams were U or M.

Table 5.2 Number of responses to Category I items by response level and by grade

Number of responses in each level or cannot be coded	Grade 6*		Grade 8*		Grade 12*	
	Adv 135	Ord 332	Adv 268	Ord 289	Adv 258	Ord 229
Prestructural	7 (5)	43 (13)	12 (4)	20 (7)	9 (3)	11 (5)
Unistructural	41 (30)	148 (45)	77 (29)	139 (47)	49 (19)	52 (23)
Multistructural	85 (63)	140 (42)	173 (65)	126 (44)	155 (61)	147 (64)
Relational	1 (1)	0 (0)	5 (2)	2 (1)	16 (6)	5 (2)
Extended abstract	0 (0)	0 (0)	1 (0)	0 (0)	24 (9)	12 (5)
Cannot be coded	1 (1)	1 (0)	0 (0)	2 (1)	5 (2)	2 (1)

* Figures in brackets represent the percentage of the responses.

The percentage of M responses are roughly the same (about 63%) for grade 12 students and grade 6 and 8 advanced school students. For grades 6, 8 and 12, within the advanced stream the percentages of responses at the M level or above were 64%, 67% and 76% respectively. Within the ordinary stream the corresponding percentages were 42%, 45% and 71%. The advanced school students' understanding levels are higher than the ordinary school students in each grade and older students tended to give more complex responses than younger students. But, the response levels for grade 6 and 8 are very close. This means that for students without any formal probability training their understanding of impossible, possible and certain does not improve with age. For grade 12 students, they gave less U responses and more R and E responses, which shows their better understanding of probability.

Overall, it appears that for category I the stream plays a role. Advanced school students answered at a higher developmental level than ordinary school students. Within each stream, there is an increase in proficiency with age but the increase is very small for the groups of students without a background in probability.

Table 5.3 Number of responses to Category II items by response level and by grade

Number of responses in each level or cannot be coded	Grade 6* Adv 150	Grade 6* Ord 369	Grade 8* Adv 303	Grade 8* Ord 324	Grade 12* Adv 291	Grade 12* Ord 261
Prestructural	2 (1)	18 (5)	4 (1)	10 (3)	2 (1)	6 (2)
Unistructural	11 (7)	24 (7)	25 (8)	29 (9)	15 (5)	17 (7)
Multistructural	120 (81)	304 (82)	244 (81)	264 (82)	192 (66)	194 (74)
Relational	15 (10)	18 (5)	26 (9)	14 (4)	74 (25)	39 (15)
Extended abstract	0 (0)	0 (0)	1 (0)	0 (0)	3 (1)	0 (0)
Cannot be coded	2 (1)	5 (1)	3 (1)	7 (2)	5 (2)	5 (2)

* Figures in brackets represent the percentage of the responses.

In category II, the data in Table 5.3 still show that advanced school students' understanding levels are higher than that of ordinary school students in each grade, but only very slightly higher. The developmental levels of grade 6 and 8 students are virtually the same, and there is little difference between advanced and ordinary students except for slightly more advanced students answering at the relational level. Grade 12 students, who had received formal instruction in probability, gave many more relational responses than the younger students. However, extended abstract responses were rarely observed even in grade 12 advanced school students.

In this category, student age, streams and backgrounds in probability did not seem play a significant role, except for responses at the R level. It should be noted the vast majority of the responses in this category were at the multistructural level, which meant that there was less variation in this category than other categories. Actually, two-thirds of the items in this category have an R ceiling level and the other third have an E ceiling level. This means that the skewed distribution towards the M level is not a result of the ceiling effect, but the students' limited knowledge of frequentist probability.

In answering category III items, evidence that advanced school students' understanding levels are higher than the ordinary school students' and the older students tended to give more complicated responses than the younger students is obtained again (see Table 5.4). For grades 6, 8 and 12, within the advanced stream, the percentages of students giving a reason at the R level or E level were 23%, 27% and 51% respectively. Within the ordinary stream the corresponding percentages were 14%, 17% and 29%.

Table 5.4 Number of responses to Category III items by response level and by grade

Number of responses in each level or cannot be coded	Grade 6*		Grade 8*		Grade 12*	
	Adv 118	Ord 307	Adv 247	Ord 264	Adv 241	Ord 212
Prestructural	0 (0)	5 (2)	2 (1)	3 (1)	1 (0)	1 (0)
Unistructural	20 (17)	91 (30)	35 (14)	49 (19)	17 (7)	14 (7)
Multistructural	56 (48)	146 (47)	131 (53)	154 (58)	92 (38)	134 (64)
Relational	24 (20)	42 (14)	62 (25)	44 (17)	101 (42)	49 (23)
Extended abstract	4 (3)	0 (0)	5 (2)	1 (0)	21 (9)	13 (6)
Cannot be coded	14 (12)	23 (7)	12 (5)	13 (5)	9 (4)	1 (0)

* Figures in brackets represent the percentage of the responses.

A similar conclusion to that indicated in category I could be drawn, namely that for students without any formal probability training their understanding of chance comparison in one-stage experiment does not improve with age. For grade 12 students, they gave less U responses and more R and E responses, which shows their better understanding of the classical definition of probability. However, the ordinary grade 12 students only operated at a slightly higher level than advanced grade 6 and grade 8 students, and gave many fewer R and E responses than the advanced grade 12 students. The grade 6 and grade 8 ordinary students responses were at a higher level in this category than in other categories.

Table 5.5 Number of responses to Category IV items by response level and by grade

Number of responses in each level or cannot be coded	Grade 6*		Grade 8*		Grade 12*	
	Adv 60	Ord 134	Adv 113	Ord 118	Adv 106	Ord 97
Prestructural	0 (0)	6 (4)	0 (0)	2 (2)	0 (0)	0 (0)
Unistructural	18 (30)	59 (44)	46 (41)	57 (47)	10 (9)	22 (23)
Multistructural	27 (44)	55 (41)	33 (29)	41 (35)	32 (30)	45 (46)
Relational	13 (22)	13 (10)	32 (28)	16 (14)	26 (25)	22 (23)
Extended abstract	1 (2)	0 (0)	1 (1)	1 (1)	35 (33)	7 (7)
Cannot be coded	1 (2)	1 (1)	1 (1)	1 (1)	3 (3)	1 (1)

* Figures in brackets represent the percentage of the responses.

Some of the results for Category IV items are similar to those already discussed in other categories. For grades 6, 8 and 12, within the advanced stream, the percentages giving a response with reasoning at the R level or E level were 24%, 29% and 58% respectively (see Table 5.5). Within the ordinary stream the corresponding percentages were 10%, 15% and 30%. Within each grade advanced students seem to be operating at a higher developmental level than their corresponding ordinary student cohort. The variation in students' understanding is very obvious in this category. The grade 12 students in advanced school did best with 58% of their responses at the R and E level. This was the one group that had a significant proportion of their responses at the E level. For the ordinary grade 12 students, they gave far fewer U responses than the younger students, which shows their better understanding of chance comparison. The ordinary grade 6 and 8 students did worst with about 50% of their responses at the P and U level.

When comparing all the four categories, category II has the least variation in distribution with most of the responses being at the M level and very few responses at the P and E level. This result might be expected given the limited experience Chinese students have with probability. Students had almost no experience in collecting and analysing data. Even in grade 12 where there has been some exposure to statistics and probability in school, a theoretical approach rather than an experimental approach is used, so their understanding of frequentist probability was only slightly better than the younger students who have had no formal experience with statistics and probability. However, in category III and IV, when students were asked to compare likelihood, the grade 12 students gave far more relational and extended abstract responses than the younger students did because they involved chance calculations in their reasoning.

In order to provide an overview of students' understanding in all four categories the above information needs to be summarised. There are several alternative approaches. One is to pool all the responses together, group them again by grade and stream, and calculate the percentages of responses at each level across all categories. For example, a total of 463 responses were observed from the advanced grade 6 students. Nine of the 463 responses (2%) were P responses. The result of this calculation is shown in Table 5.6.

In terms of developmental responses it shows that the ordinary school grade 6 and 8 students had the poorest understanding of probability, and advanced school grade 6 and 8 students had a slightly better understanding, but not as good as ordinary school grade 12 students. The advanced school grade 12 students had the best understanding of probability and they gave more relational and extended abstract responses than other students.

Table 5.6 Number of responses to all the four categories' items by response level and by grade *

Percentages of responses in each level or cannot be coded	Grade 6		Grade 8		Grade 12	
	Adv 463	Ord 1142	Adv 931	Ord 995	Adv 896	Ord 799
Prestructural	9 (2)	72 (6)	18 (2)	35 (4)	12(1)	18 (2)
Unistructural	90 (19)	322 (28)	183 (20)	274 (28)	91 (10)	105 (13)
Multistructural	288 (63)	645 (57)	581 (62)	585 (58)	471 (54)	520 (66)
Relational	53 (11)	73 (6)	125 (13)	76 (8)	217 (24)	115 (14)
Extended abstract	5 (1)	0 (0)	8 (1)	2 (0)	83 (9)	32 (4)
Cannot be coded	18 (4)	30 (3)	16 (2)	23 (2)	22 (2)	9 (1)

* Figures in brackets represent the percentage of the responses. More than 96% of the responses collected in each grade and each stream were coded. The lowest value 96% was observed from advanced grade 6 students.

If we ignore the stream and compare the results for grades 6 and 8, the two grades without any formal probability training are virtually identical. It appears that there is no improvement in developmental level at the lower grades. A change in sophistication does not come about just because the students have two more years' schooling or are two years older. Compared to the grade 6 and 8 students, the grade 12 students had approximately the same percentage of multistructural responses but a much higher percentage of relational and extended abstract responses and a much lower percentages of prestructural and unistructural responses. While age might be a factor in increasing grade 12 students' response level, it seems reasonable to conclude that the improvement could be attributed in part to exposure to the teaching of probability in school. This conclusion is reinforced by the lack of increased response levels between grades 6 and 8 and the results of the teaching intervention in this study.

The numerical index

A second method is to use a Rasch partial credit model that enables an analysis to be undertaken when all students did not write all items. It can be applied in situations in which performances on items are recorded in two or more ordered categories and there is an intention to combine results across items. It has the following property that not all persons need to have been given all the items, providing there is an overlap (Andrich, Hess & Ryan, 1998). The parallel design of the nine sets of questionnaires in this study meets the overlap requirement. For this reason, the model and the programme Rasch Unidimensional Measurement Models (RUMM Laboratory Pty Ltd) was selected to assign an overall numerical understanding location to each student.

The analysis shows that except for one grade 6 student, d3064171m, who gave P responses to all the items and was identified as an extreme student (his location was -10.136), all the other 566 students were located in the interval (-2, +5). The mean for all the students' location values was 0.98 and the SD was 1.26. In order to discuss the data the interval was divided into five equal segments and the students were grouped according to their locations. This information is contained in Table 5.7 and is represented diagrammatically in Figure 5.2.

Table 5.7 Overall picture of students' numerical understanding location index

Number of students located in each interval	Grade 6*		Grade 8*		Grade 12*	
	Adv 50	Ord 124	Adv 101	Ord 108	Adv 97	Ord 87
(-11, -1.6]**	0 (0)	3 (2)	1 (1)	0 (0)	0 (0)	0(0)
(-1.6, -0.2]	6 (12)	44 (35)	11 (11)	29 (27)	3 (3)	6 (7)
(-0.2, 1.2]	28 (56)	61 (50)	47 (46)	59 (54)	20 (21)	34 (39)
(1.2, 3.6]	16 (32)	16 (13)	42 (42)	20 (19)	55 (56)	42 (48)
(3.6, 5)	0 (0)	0 (0)	0 (0)	0 (0)	19 (20)	5 (6)

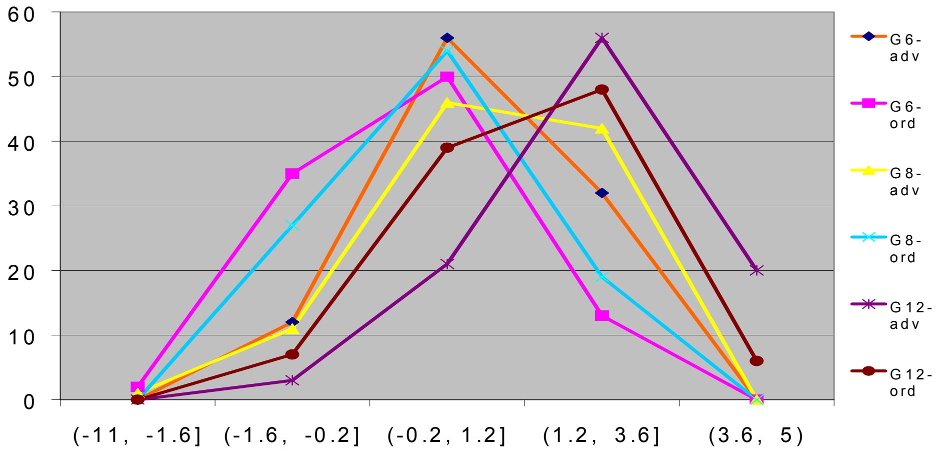
* Figures in brackets represent the percentage of students whose location at each interval for the corresponding cohort. The higher the location the better understanding.

** On the surface the first interval looks much larger than the other four. Only one of the 567 students, an outlier, was outside the interval (-2, +5). This interval was then divided into 5 equal segments. When this student was included for completeness is appears that the first interval is much larger than the others.

The data reinforce the earlier conclusion that ordinary school grade 6 and 8 students had the poorest understanding of probability, advanced school grade 6 and 8 students had a slightly better understanding, but not as good as ordinary school grade 12 students. Advanced school grade 12 students had the best understanding of probability among all the students.

This conclusion has been obtained from Table 5.6. However, since each student now has a numerical understanding location it makes the difference between persons, grades or streams easier to observe. From Figure 5.2, it seems that within the same stream, grade 8 students' understanding was slightly better than that of grade 6.

Figure 5.2 Overall picture of students' numerical understanding location index



It should be noted that the five subintervals do not correspond to the five SOLO levels. Since not all the students completed the same questionnaire, it also does not mean that a student who gave more R and E responses in total must be located at a higher numerical position, although usually the order of the two indices were in accord. For example, two grade 12 students, i7123520 and i7124300, in two classes of an ordinary school obtained the highest location 4.819. Their identical response levels on each item are shown below. They achieved the ceiling level in seven items.

Item	I2(4)	I3(3)	II1(7)	II2(2)	II3(6)	III4(3)	III4(4)	IV4	IV5
Observed	M	M	M	R	M	R	R	E	E
Ceiling	M	M	R	R	E	R	R	E	E

Another grade 12 student, h5121380, in an advanced school, whose location was 4.239, gave the following responses to each item of another questionnaire. Three M's, four R's and two E's responses in all. She achieved the ceiling level on six items.

Item	I2(6)	I3(2)	II1(2)	II2(6)	II3(7)	III3(1)	III3(2)	IV2	IV10
Observed	M	M	R	R	M	R	R	E	E
Ceiling	R	M	R	E	R	R	R	E	E

In looking at the SOLO levels, the latter student gave one more R response in total, but considering ceiling levels, her location was slightly lower than the previous two students.

Although the location does not correspond exactly to the total SOLO score (assigning 0, 1, 2, 3 and 4 to P, U, M, R and E responses respectively), students with higher location indices generally gave responses at the higher levels. This is illustrated by the examples above where all the three students gave a substantial number of R and E responses. The one student who was an extreme student at the lowest end gave prestructural responses on all items. It seems reasonable to conclude that this analysis gives a realistic picture of the situation and is consistent with descriptive analyses outlined earlier in this chapter.

Other

The response level analysis was based on the students' written responses. It could be argued that if students had been given the opportunity to give a verbal response the levels might have been different, and more often higher. However, the interview data showed most students did not change their level of response. The changes that did occur were in three situations. First, when the interviewees had alternative strategies for an item but they were not sure which one was better, when they were questioned they changed to the alternatives and sometimes their SOLO level changed. Second, when the interviewees were asked further questions such as, "Could you tell me the probability of the outcome happening?", sometimes more sophisticated responses were observed resulting in a higher SOLO level.

Third, for the students who gave blank or incomplete answers or misread problems in a questionnaire, when they were required to answer the items again, a change occurred. However, for most of the students in the third situation their answers were still at a low level.

In conclusion, the interviews supported the conclusion that the written responses are a valid reflection of the developmental level of the students on that item.

Summary

As a result of this study, the following cognitive framework guided by SOLO taxonomy was generated (see Table 5.8). It indicates the developmental process of students' understanding of probability.

Table 5.8 A cognitive framework relating to different categories of items used in this study

SOLO Level	Summary Description
P	Blank, fully irrelevant, illogical, egotistic answers or inability to become engaged in item answers.
U	Explains that probability just means may or may not happen and believes chance cannot be measured mathematically so chance comparison is impossible. Considers an incomplete set of outcomes in solving problem.
M	May consider all possible outcomes for a one-stage and sometimes for a two-stage experiment in qualifying uncertainty or estimating subjective chance value. For example, assigns an equal chance to each possible outcome for fairness. Interprets most likely to happen as meaning it should happen or interprets chance by frequency but without fully understanding the role of repetition. Uses rudimentary non-proportional reasoning in chance comparison.

R	Groups all possible outcomes in favour of a target event together and uses ratio as a measure of probability. Uses proportional reasoning in chance comparison. Knows that a larger number of repetitions is a more reliable predictor and expresses the idea of making a few repetitions automatically.
E	Assigns a calculated probability value in complicated situations, for example, involving two bags, two spinners and bases chance comparison on the values. Uses a generative strategy to construct sample space in a two- or three-stage experiment to work out probability. Suggests collecting data from a series of experiments and finding trends across sampling.

The data presented in this chapter show that the developmental level for grade 6 and 8 students is lower than that for grade 12 and it appears that there is no obvious improvement from grade 6 to grade 8. Also, the lack of improvement from grade 6 to grade 8 implies that for students in these grades two years of schooling does not play a role in developing their understanding when there is no instruction in probability. The grade 12 students are the only students who have had formal instruction in probability and, although there were still many lower level responses, they generally operate at a higher level. It would appear that even a short unit on probability has an impact on students' understanding of probability. It is also possible that the additional maturity from grade 8 to grade 12 has some impact. Advanced students at a given grade have a higher developmental level than their corresponding ordinary student cohort.

As indicated at the beginning of this chapter the SOLO framework was used for analysis rather than that developed by Piaget and Inhelder (1975). Piaget and Inhelder identified three major stages in children's development of probability and they believed that there was a relationship between children's ages and the three stages. In contrast, the SOLO taxonomy labels specific responses according to level, independent of age. However, both Piaget and Inhelder and this author tried to analyse features of the understanding by using a hierarchical sequence levels. A child identified by Piaget and Inhelder as being at the first stage would probably most often give P or U levels responses as described in Table 5.8. A child in the second stage could probably give many more M or R levels' answers and a child in the third stage might be able to give E responses.

Chapter 6 Teaching Intervention

The results presented in both chapters 4 and 5 show that many students in each grade and each stream had a poor understanding of frequentist probability. Most of the students in this study had applied the following four misconceptions (which are related to the frequentist definition of probability) at least once when answering the items on the questionnaire. They are: the outcome approach; one trial is unrelated to other trials; interpreting chance by data matching or words matching and increasing repetition is not better for predicting. This result indicates that the understanding of using frequencies as an estimate of probability is not developed naturally. Particularly, in light of the results for the grade 12 students where, although they had studied some introductory and largely classical probability in their mathematics course, there were still a substantial number of students who continue to have misconceptions associated with frequentist probability. The purpose of this teaching intervention was to determine whether an activity-based, short-term teaching programme could help ordinary school grade 8 students understand probability better and overcome some of their misconceptions.

The information on the design of teaching intervention, selecting of the subjects and the general process for the teaching intervention were described in chapter 3. In this chapter, the results of the teaching intervention are presented. Before discussing the results an outline of the six lessons is given.

Outline of the Lessons

A series of six lessons was given to two grade 8 classes in an ordinary school in Shanghai over a three-week period. During the teaching intervention students did activities such as carrying out experiments, recording outcomes, observing the relative frequencies and discussing questions using the data.

The structure of each day's lesson is similar, as illustrated below:

1. Introduction or review.

2. The main problem being presented was described. The main problems for each day are listed below.

Day 1: You are given an ordinary six-sided die. For each roll, the number you will roll depends on luck. However, is there any pattern that exists behind the random phenomena?

Day 2: When you roll an ordinary six-sided die three times, do you think the chance that the number 6 is rolled in each trial is the same or the chance is changed based on past results?

Day 3: Flip a fair coin twice. You may get two heads, two tails or one head and one tail. Which of the outcomes has the greatest chance, or do they have the same chance, or is it impossible for us to decide? What is your opinion? How about the situation of flipping a coin three times?

Day 4: You are given two different size spinners. Spin each spinner's arrowhead with all your strength. Suppose you want the arrowhead to stop in the blue part. Which spinner do you think is more likely to stop in the blue part?

Day 5: Here are 20 playing cards. However, we don't know how many cards are red and how many cards are black. Can we make a smart guess without turning over all the cards? What's your plan?

Day 6: In a test, suppose you had no idea for three True or False items and you decided to guess the answers. The chance that you got the correct answer for each item is 50%. Please find out the chance that you would get two or more correct answers out of the three items. Tell me your plan.

3. An explanation was given on how to carry out the activity and how to record the data.

4. A presentation outlining some of the wrong conclusions relating to the specific task that were observed in the main study, such as number 6 is the hardest number to be rolled, was made and the wrong conclusions were briefly discussed.

5. Students did the activity in pairs and recorded the results.
6. The students' data were pooled or they were given the computer-generated data, depending on time.
7. The wrong conclusions were discussed again in light of the results of the experiment they had completed.

Each student was given a workbook that described the daily lesson. The workbook gave the main teaching points on a day-to-day basis, how to undertake activities, how to enter the experimental data into tables, the questions for whole class discussion and an appendix with data generated by computer simulations. The materials for the first day's lesson are included in Appendix D.

Teaching Intervention Results

All the students whose data were analysed were tested and interviewed both prior to and after the teaching intervention. Two students in the class without a computer were eliminated from the analysis because they either missed the post-test or missed the interview after the post-test due to illness. The items used in the pre-test and post-test were the same or parallel versions. Most of them had been used in the main study. Before reporting the results, several points should be noted since it is likely that each of them could have influenced the students' responses.

First, according to the curriculum standards, data collection and data description will be taught in grade 9. So the grade 8 students had virtually no experience in school of dealing with real data prior to the teaching intervention. The students limited knowledge of recording data and interpreting data might have restricted their learning of probability.

The second point concerns the teaching time. For the class without a computer, half of their classes were arranged in self-study periods after school (students usually do their homework during such periods), so a few students did not fully focus on the learning and tried to do their homework at the same time. For the class with the computer, students often had a bit less teaching time since the computer lab was in another building.

Finally, the teaching intervention was very short, consisting of only six lessons within a three-week period. Also, the topic was unrelated to their formal curriculum. These three points all have the potential to have a negative effect on the students' achievement.

However, taking part in an educational experiment was a new experience for the students. Involvement in an activity-based new teaching programme with an unfamiliar teacher might result in a greater effort from some of the students and have a positive effect.

The results of the teaching intervention are reported under the four categories that this study investigated. For each item, the data are organised in three parts: the answer, SOLO level and reason. Since each student was interviewed the data combined information from both the questionnaire and the interview. When a written response on an item was unclear for coding, either in terms of SOLO level or reason, further questions were asked in the interviews.

In the following tables, the symbol ✔ stands for correct answers or reasons, and ✕ stands for incorrect answers or reasons. W/ stands for the class with the computer (26 students) and W/O stands for the class without a computer (23 students). The letters J and K refer to the questionnaires used before and after the teaching intervention respectively. The number following each letter refers to the order of the item in the questionnaire, for example, J1 is the first item used in the pre-test. The letter G stands for misconception Group; for example, G1 means the misconception belongs to the Group 1 misconceptions reported in chapter 4.

Since in chapters 4 and 5 there were many detailed examples describing each group of misconceptions and each of the SOLO levels, in this chapter only a few examples of the misconceptions and levels are used to illustrate the change observed between the two tests.

Category I items

The first two items in the two questionnaires asked the students to identify impossible, possible and certain events. The situation for items J1 and K1 dealt with rolling dice and J2 and K2 with drawing card(s). The correct choice for J1 and K1 is possible and that for J2 and K2 is certain.

Table 6.1 shows most of the students were able to correctly identify the three kinds of events in both the pre and post test.

Change in SOLO levels

Student j8081241 gave a P response for J1. He selected both impossible and possible as his answer giving the reason that the result depends wholly on luck. In the interview, he explained that he chose the two options to mean sometimes it is impossible and sometimes it is possible. In the post-test, he gave an M response for K1. He correctly chose possible and indicated that for each die the probability of rolling an even number is 50%.

Table 6.1 Data on students' responses to impossible, possible and certain events items

Responses		J1: all three numbers rolled are odd numbers		K1: all three numbers rolled are even numbers		J2: all three cards drawn are smaller than a Jack		K2: the card drawn is smaller than a Jack	
		W/	W/O	W/	W/O	W/	W/O	W/	W/O
Answers	✔	22	17	26	23	21	21	25	22
	✗	4	6	0	0	5	2	1	1
SOLO levels	P	1	4						
	U	19	16	1		5	3		1
	M	6	2	22	21	21	20	26	22
	R		1		1				
	E			3	1				
Reasons	✔	22	17	24	19	18	20	26	22
	✗	G2:2	G2 G6:1	G14:2	G2, G14:1	G2 G5:1	G2:3		G2:1
	✗	G6:2	G2:3		G14:2	G2:7			
	✗		G3:1		G4:1				
	✗		G6:1						

G2: example-based interpretations for possible and impossible
G3: possible means certain
G4: chance cannot be measured mathematically
G5: equiprobability
G6: outcome approach
G14: used own methods in chance calculation

Another student, j8082201, gave the only R response for item J1. He listed four possibilities (odd-odd-even; odd-even-even; odd-odd-odd; even-even-even) when three dice were rolled. Then he made a subjective estimation for the chance that three odd numbers were rolled: smaller than 30%, about 24.6%. In the post-test, he gave an E level response for K1 as he listed all the eight possibilities and indicated that the probability of rolling three even numbers was 1/8.

Overall, the data show that the SOLO level of the students' responses improved after the teaching intervention. This was particularly noticeable for items J1 and K1 where the majority of responses went from a U level response to an M level response.

<u>Change in reasoning</u>

For items J1 and K1, 8 of the 10 students who used misconceptions in the pre-test did not use them in the post-test. The responses of these two students are discussed later. In answering K1, five students used their own methods in chance calculation (G14). One student, tried to calculate the chance value of three even numbers are rolled in his written answer, four other students used G14 during their interviews. They were asked to compare the chance of rolling an even number with one die with the chance of rolling three even numbers with three dice. They said that both situations have a 50% chance. Since they did not use equiprobability as their reason, these responses were coded as using G14. This error was probably due to the fact that during the teaching intervention the calculation of probability was discussed and these students had not fully comprehended how to calculate probability correctly in such a complex situation.

For items J2 and K2, all 11 students who used misconceptions in the pre-test did not use them in the post-test.

The responses of two students who continued to uses misconceptions are now discussed. The first student j8082150 gave a correct answer for J1 but without giving any reason. In answering K1, she still answered correctly that all three numbers rolled are even numbers was possible. Her reason was that *we couldn't predict the possibility of each trial,* which was coded as using the misconception G4 (chance cannot be measured mathematically).

In answering K2, she gave the only wrong response. She chose *possible* and explained that *we cannot predict the outcome of each trial.* In the interview, she said that her statement that one *cannot predict the* *possibility of each trial* in K1 has the same meaning as saying that one *cannot predict the outcome of each trial* in K2. She said that *all the cards that are left are from 1 to 10, so drawing a card smaller than a Jack is possible.* Her response for K2 was applying misconception G2 (example-based interpretations for possible and impossible).

The other student, j8082110, was coded as consistently applying misconception G2 in J1 and K1. Actually, the misconception was observed in her interviews when she was asked to answer other questions that were not on the written test. In the two interviews, she was asked some common questions (sequentially), such as whether it is impossible, possible or certain that the number 6 is rolled. How about the number rolled is from 1 to 5 or 1 to 6? She answered *possible* to the event 1-6 in both interviews, applying G2. But in the second interview, when she was asked if she could find the probabilities of these events, she said she could and gave the probabilities as 1/6, 5/6 and 1. At that time, she found her mistake herself and corrected her answer to *certain.*

The data in this category suggests that virtually all the students could correctly identify impossible, possible and certain events after being taught the meaning of the three words. However, after the teaching intervention some of the students attempted to calculate probability in a three-stage problem and this resulted in a new misconception associated with incorrectly calculating a probability.

Category II items

The next three items in the two questionnaires related to the frequentist definition of probability, which was the main focus of the teaching intervention.

Among the six items, J4 and K4 are parallel to each other (the context is changed) and the other two pairs are unchanged.

J3 (the same as K3, and this was item II2(2) in the main study):

> A mathematician filled a bag with black and white marbles. He didn't really know how many black marbles and white marbles were in the bag. After mixing them thoroughly, he took a look and predicted that "if I pull out a marble from the bag without looking, the chance that it will happen to be white is 50%." Which of the following has the closest meaning to "the chance that it will happen to be white is 50%"?
>
> a) the marble pulled out may be white or may be black. The mathematician doesn't really know what the result will be
> b) suppose that the game is repeated 10 times, white marbles are pulled out around 5 times
> c) suppose that the game is repeated 100 times, white marbles are pulled out exactly 50 times
> d) suppose that the game is repeated 100 times, white marbles are pulled out around 50 times

J4 (item II3(6) in the main study):

> A weather forecaster said that "tomorrow there is a 30% chance that it will rain." The next day it doesn't rain. Do you think the forecaster's prediction is accurate or not?

K4 (item II1(6) in the main study):

> Before the final match, the coach predicted that "based on the information I have, our chance of winning the game is 30%." The game is played and the team loses. Do you think the coach's prediction is accurate or not?

J5 (the same as K5):

> Medical research indicated that about 1 in 20 people develop a particular illness. However, in a sample of 322 people, it was found that 39 had developed this disease. Are these two pieces of information in conflict? If you really want to get a clearer picture of the real incidence of this disease, what will you do? Why?

Since the items J4, K4, J5 and K5 are open-ended problems, it is inappropriate to simply label a response as correct or incorrect based solely on a student's conclusion. This can be illustrated by the following response for item J4. The expected answer for J4 is that at present it cannot be decided or a similar statement. But consider the following answer that *the* *weather forecaster's prediction was accurate, because 30% chance of rain didn't mean it should rain, the prediction was not proved to be wrong*. We cannot really say that the response is correct, but it is reasonable and cannot be coded as incorrect. Given that coding an answer as correct or incorrect for the open-ended items could be misleading, the first part of Table 6.2 only includes correct or incorrect information on the answers for the two multiple-choice items.

Table 6.2 shows that students in both classes improved considerably in terms of the number of correct answers (K3), SOLO levels and reasons between the two tests when answering items K3 and K5. However, there was little improvement when answering K4 and it seems that the outcome approach (G6) was very stable.

<u>*Change in SOLO levels*</u>

The following two examples show that the students had a better understanding of frequentist probability after the teaching intervention.

For item J3, student j8082130 thought that option (a) was the best answer as she believed that it is impossible for the mathematician to make as accurate a prediction as indicated by options (b), (c) and (d). Her response was coded as a U level response since she interpreted a chance value between 0 and 1 as the target outcome may or may not happen. In the post-test, she selected the correct answer, option (d), and explained

> *It's impossible to predict exactly how many times he can pull out white marbles, but it's possible to predict the probability is 50%. Option (b) is not so good as it involves fewer repetitions. Comparatively, option (d) involves more repetitions.*

Her second response was coded as an R level response since it contains the elements of abstract thinking that the more repetitions the more reliable the estimation.

Table 6.2 Data on students' responses to interpretation of chance values items

Responses		J3 (K3): the closest meaning to "the chance of pulling out a white marble is 50% chance"			
		J3		K3	
		W/	W/O	W/	W/O
Answers	✔	8	3	21	18
	X	18	20	5	5
SOLO levels	P	0			
	U	10	10	2	1
	M	10	12	3	3
	R	6	1	21	19
Reasons	✔	20	13	23	21
	X	G1:1	G4:2	G4:1	G8:1
	X	G7:1	G8:4	G9:2	G9:1
	X	G8:2	G9:4		
	X	G9:2			

Responses		J4: 30% chance of rain, no rain, accurate?		K4: 30% chance of win, lose, accurate?		J5 (K5): 1 in 20 people has the illness, 39 in 322 people have the illness, in conflict? How to know the real incidence?			
						J5		K5	
		W/	W/O	W/	W/O	W/	W/O	W/	W/O
SOLO levels	P		1						
	U	1	1	1		8	3	2	
	M	25	21	24	21	13	15	4	7
	R			1	2	5	5	20	14
	E								2
Reasons	✔	12	10	17	13	23	21	24	22
	X	G6:14	G1:1	G4:1	G6:10	G9:1	G9:1	G9:2	G9:1
	X		G6:12	G6:8		G10:2	G9, G10:1		

G1: subjective judgements
G4: chance cannot be measured mathematically
G6: outcome approach
G7: one trial is unrelated to other trials
G8: interpreting chance by data matching or word matching
G9: increasing repetition is not better for predicting
G10: positive and negative recency

For item J5, student j8082120 gave an R level response. She believed the two pieces of information were not in conflict as the real situation could be not as the same as the theory, and she planned to average the two data, make one more survey, then the real answer might be found. In the interview, she said if the data were still not close to each other, she preferred to average them again. In the post-test, she still believed that the two pieces of information were not in conflict. She wrote

> *I will do more survey, until the last result I get is quite close to the previous one. I believe it will tally the theoretical value. Only if you repeat the experiment more times, your result could be close to theoretical value and could be more accurate.*

When she was asked what she would do if the result obtained from a sample of one thousand people was neither 1/20 nor 39/322, she said she would survey another thousand people, enlarge the sample, until the fluctuation became less and less. Her second response was coded as an E level response since she suggested making a large number of repetitions to make a conclusion.

Only one student gave a lower level response than before. This will be illustrated in the next section.

A final note on the change of SOLO levels between the two tests. The data also indicate that the J4/K4 pair was the only one where the students' SOLO levels showed very little improvement. A possible reason is that even the students who did not use the outcome approach and answered correctly, if they did not suggest making more trials their responses still could not be coded as higher than the M level. This is illustrated in the following example. In answering J4, student j8081231 gave an M level response using the outcome approach:

> *Accurate. Since the weather forecaster said the chance that it will rain is 30%, the chance that it will not rain should be 70%. 70%>30%, it will not rain.*

When he was asked about the situation if the percentage was changed from 30% to 80%, he said that *the prediction is inaccurate.* He changed his judgement along with the change of the percentage involved in the prediction. In answering K4, he also gave an M level response*, because the chance of winning the game is between 1% to 100%.* When he was asked what would happen if the percentage was changed from 30% to 80% he said that *the prediction is still accurate.* The outcome approach had disappeared but he made no suggestion of taking more observations.

<u>*Change in reasoning*</u>

For item J3 (K3), only three of the 16 students who used misconceptions in the pre-test used them in the post-test. The situation for item J5 (K5) is similar. All five students who used misconceptions in J5 gave the correct responses in K5. However, five wrong reasons were observed in K3 and three in K5. In the following examples, one shows a student who kept a misconception and the other shows a new misconception that arose between the two tests.

Student j8081020 used the misconception that increased repetition is not better for predicting (G9) in answering both J3 and K3. She chose both options (b) and (d) as her answer to J3 since she believed *there was no difference between 10 times and 100 times.* So her reason for J3 was coded as using G9. In the post-test, she correctly selected option (d) with a correct reason. But in the interview, she was asked if another option (e) was added, whether it would change her answer.

e) suppose that the game is repeated 150 times, white marbles are pulled out around 75 times

She said she would still chose option (d). It indicated that she still retained the misconception G9. In answering item K5, she suggested surveying 100 people. When she was asked why she decided to sample 100 people, she said because one hundred made calculation easier. When asked about surveying 1000 people, she said it is the same as surveying 100 people. She used G9 again.

Student j8082050 gave a correct R level response for J3 but used the misconception of interpreting chance by word matching (G8) for K3. She wrongly chose option (c) for K3, because *the mathematician said **just**, so his original intention should be **exact** not **around**. …* In the interview, she still chose (c) but also expressed her hesitation in choosing (c) *because based on my experience the option (d) should be the best answer*. However, she made a mistake in K3 because she overemphasised the word *just* in reading the item. For category II items, this was the only one response that resulted in a lower level than before.

For items J4 and K4, the data show that 11 out of 26 students who used the outcome approach for J4 answered K4 correctly. However, the outcome approach (G6) was still one of the main misconceptions observed after the teaching intervention. One tentative reason is that the outcome approach is not easy to overcome with a short teaching intervention, a possibility has been suggested in the literature (for example, Konold 1989).

Category III items

Items J6/K6, J7/K7 and J9/K9 in the two questionnaires are related to chance comparison in one-stage experiments.

J6: A school has 500 girls and 550 boys in it. Each pupil's name is written on a piece of paper and all the names are put into a box and mixed thoroughly. The principal picks 1 name out of the box casually without looking. Which statement below is correct?

a) it is more likely to pick out a boy's than a girl's name
b) it is less likely to pick out a boy's than a girl's name
c) it is equally likely to pick out a boy's as a girl's name
d) it is impossible to compare the likelihood of the two outcomes

K6: (item III3(1) in the main study).

The only change from J6 is the school has 400 girls and 440 boys.

J7: There are 9 red marbles and 18 black marbles in bag A. There are 60 red marbles and 80 black marbles in bag B. Mix the marbles in each bag thoroughly. Close your eyes and suppose you want to pull out a black marble. Which statement below is correct?

a) the likelihood of pulling out a black marble from bag A is greater than that from bag B
b) the likelihood of pulling out a black marble from bag A is less than that from bag B
c) the likelihood of pulling out a black marble from bag A is the same as that from bag B
d) it is impossible to compare the likelihood of the two outcomes

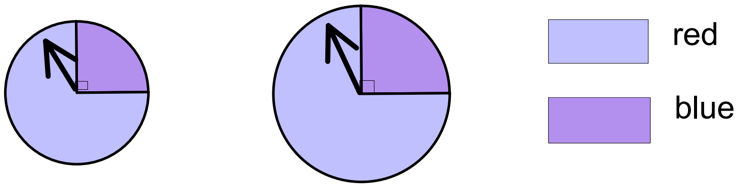
K7: (item III5 in the main study).

The only change from J7 is that bag A contains 8 red marbles and 16 black marbles and that bag B contains 50 red marbles and 70 black marbles.

J9: (item III9 in the main study).

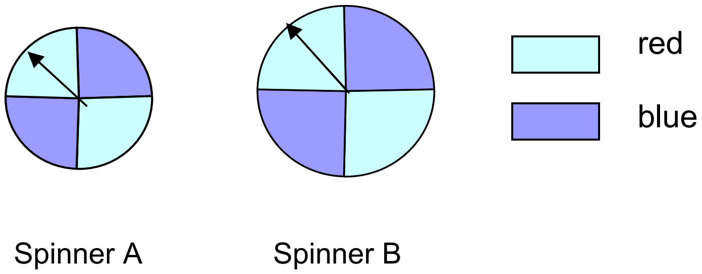
Spin each spinner's arrowhead with all your strength. Suppose you want the arrowhead to stop in the blue part. Which statement below is correct?

a) the arrowhead of spinner A is more likely to stop in the blue part than that of spinner B
b) the arrowhead of spinner A is less likely to stop in the blue part than that of spinner B
c) the arrowhead of spinner A is equally likely to stop in the blue part as that of spinner B
d) it is impossible to compare the likelihood of the two outcomes



spinner A   spinner B

K9: the only change with J9 is the two spinners



The data in Table 6.3 show again that generally there is a clear improvement in both classes in students' answers, SOLO levels and reasons between the two tests, especially for items K6 and K9. However, since the teaching intervention was focused on the frequentist definition of probability not the classical definition of probability, the data show that many students still did not know the correct method for calculating chance. They used their own methods of chance comparison (G11) or chance calculation (G14) in the two complex one-stage experiments (K7 and K9).

Change in SOLO levels

In this category, only two students gave a lower level response in the post-test than they did in the pre-test. Student j8082110 was such a student. She gave an R level response for J7 as she found that the number of black marbles was double the number of red marbles in bag A, she correctly indicated that it is more likely to pull out a black marble from bag A. However, in answering K7, she thought there were more black marbles in both bags, therefore, she concluded that it was equally likely to pull out a black marble from each bag. This response was coded as an M level response, using equiprobability (G5). In the interview, when she was shown her own written answers in the two tests and was asked to make the final decision, she still chose (c), but could not explain why (a) was not as good as (c).

Table 6.3 Data on students' responses to one-stage items

Responses		Draw a slip from a bag				Choose a bag, a black marble is wanted				Choose a spinner, want the arrow to stop in blue			
		J6		K6		J7		K7		J9		K9	
		W/	W/O	W/	W/O	W/	W/O	W/	W/O	W/	W/O	W/	W/O
Answers	✔	14	11	25	20	12	13	19	17	11	14	20	21
	×	12	12	1	3	14	10	7	6	15	9	6	2
SOLO levels	P												
	U	5	8		1	5	5		1	10	9		
	M	21	15	17	13	13	6	7	5	13	6	7	3
	R			9	9	8	12	11	10		8	4	2
	E							8	7	3		15	18
Reasons	✔	12	10	25	19	7	11	15	13	5	10	18	21
	×	G4:6	G4:9	G14:1	G4:1	G4:6	G4:4	G5:3	G4:1	G1:1	G1:3	G1:1	G5:1
	×	G5:7	G5:4		G5:2	G5:7	G4, G11:1	G11:5	G5, G14:1	G4:8	G1, G4:1	G4:1	G11:1
	×	G14:1			G14:1	G11:5	G5:2	G14:3	G5:3	G4, G5:1	G4:7	G11:6	
	×					G14:1	G11:4		G11:2	G5:5	G5:1		
	×						G14:1		G14:3	G11:6	G11:1		

G1: subjective judgements
G4: chance cannot be measured mathematically
G5: equiprobability
G11: used own methods in chance comparison
G14: used own methods of chance calculation

The following two examples show that for some students their knowledge of chance comparison has improved.

In answering J7, student j8082040 selected option (d) and said the outcome *depends wholly on luck so it is impossible to indicate which is the most likely one*. The response was labelled as a U level response using the misconception that chance cannot be measured mathematically (G4). In answering K7, she correctly chose option (a) and based the chance comparison on working out the probabilities of pulling out a black marble from each bag. Her second response was labelled as an E level response.

The SOLO level change in student j8082060 responses for J9 and K9 is also clear. In answering J9, she noticed that spinner B (the larger spinner) has both larger blue and red areas than spinner A (the smaller spinner), and she felt that option (c) should be the correct answer, an M level response. In answering K9, she indicated that the ratio between red area and blue area on each spinner is equal to ½, so the chances for the two spinners are the same, ½. In the interview, when she was required to answer item J9 again, she said she still chose the option (c) but she would like to change the reason to *the two spinners have the same chance, ¼, to let their arrowheads stop in the blue part*. It indicates her response to the item was at the E level.

<u>Change in reasoning</u>

There are four possibilities that could happen in students' reasoning between the two tests: (correct, correct), (correct, incorrect), (incorrect, correct) and (incorrect, incorrect). The first three situations have been illustrated by examples shown in the last sections. Here are two examples of where both reasons were incorrect. One example is where students used the same misconception in both and the other is where they changed from using one misconception to another misconception.

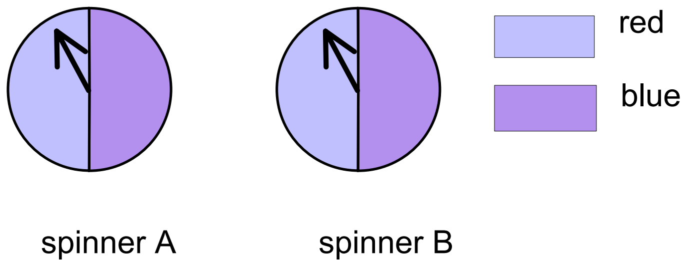
For item J7, student j8081221 thought it was more likely to pull out a black marble from bag A since *there're fewer marbles in the bag*. This reason was coded as using the misconception G11 (used own methods in chance comparison). For item K7, he changed his answer to bag B since *there're* *more marbles in the bag* (still coded as using G11). In the interview, when required to explain the change in choice, from the bag with the smaller total to the bag with the bigger total, he said it was because he asked his father for the answer to J7 after the pre-test and his father told him the more marbles the easier it is to pull out what you want.

Item J9 and K9 are the final items that appeared on the two tests. In the interview, student J8081171 said that for item J9 it was very difficult to predict the final positions of the arrowheads. Therefore, he chose option (d) (it is impossible to compare the likelihood of the two outcomes). This response was coded as using the misconception G4. For item K9, he chose the bigger size spinner as he thought the blue area on the larger spinner was bigger than that on the smaller spinner. His second response was coded as using the G11.

Category IV items

Items J8 and K8 are related to chance comparison in two-stage experiments. They were adapted from items IV2 and IV4 used in the main study. As was indicated in chapter 4, since there was no equiprobable option for the two original items, it was difficult to identify whether students used the reason that chance cannot be measured mathematically or used equiprobability when they chose option (d) it is impossible to indicate which one is the most likely among the above three outcomes, but did not explain the answer clearly. For the teaching intervention, option (d) was separated into two options (d) and (e).

J8: (adapted from item IV4 in the main study).



Spin each spinner's arrowhead with all your strength. Which statement below is correct?

a) it is most likely that both the arrowheads stop in the red part
b) it is most likely that both the arrowheads stop in the blue part
c) it is most likely that one arrowhead stops in the red part and the other arrowhead in the blue part
d) the above three outcomes have the same likelihood
e) it is impossible to measure the likelihood of the three outcomes

K8: (adapted form item IV2 in the main study).

There are 2 white marbles and 2 black marbles in bag A. There are also 2 white marbles and 2 black marbles in bag B. Mix the marbles in each bag thoroughly. Put your hands in the two bags and pull out a marble from each bag without looking. Which statement below is correct?

a) it is most likely that both marbles are white
b) it is most likely that both marbles are black
c) it is most likely that one marble is white and the other one is black
d) the above three outcomes have the same likelihood
e) it is impossible to measure the likelihood of the three outcomes

Table 6.4 shows again that generally there is a clear improvement in both classes in students' answers, SOLO levels and reasons between the two tests, especially in the class without a computer. However, a large number of students in the two classes used misconceptions in answering K8.

Table 6.4 Data on students' responses to two-stage items

Responses		J8: spin two arrowheads, compare chances of all possibilities		K8: pull out a marble from each bag, compare chances of all possibilities	
		W/	W/O	W/	W/O
Answers	✔	3	5	10	17
	✕	23	18	16	6
SOLO levels	P				
	U	14	9	1	
	M	12	13	18	7
	R		1	3	5
	E			4	11
Reasons	✔	2	4	8	15
	✕	G1:2	G1:2	G4:4	G5:5
	✕	G4:11	G1,G5:2	G4,G11:1	G11:3
	✕	G5:9	G4:7	G5:11	
	✕	G4,G5:2	G5:6	G11:1	
	✕		G11:2	G14:1	

G1: subjective judgements
G4: chance cannot be measured mathematically
G5: equiprobability
G11: used own methods in chance comparison
G14: used own methods of chance calculation

Change in SOLO levels

All the students gave a higher or an equal level response for K8 than for J8. The change in level of response is greater in the class without a computer than with the computer.

An example of the change in response level is illustrated below. In answering J8, student j8082201 selected option (e) that *it is impossible to measure the likelihood of the three outcomes,* with his reason being that *each outcome was possible*. The reason was coded as a U level response using the misconception of chance cannot be measured mathematically (G4). In answering K8, he selected option (c), and gave a correct E level reason:

> *There are the following possibilities: white and white, white and black, black and black, black and white. Each of them has the probability ¼. White and black = black and white, so the probability for one marble is white and the other one is black is ½. 1/2 >1/4.*

When he was asked to solve J8 again, he said *the two items are the same, option (c) is correct.*

The main reason for such a large improvement in this item is that the item is similar to the activity on Day 3. On that day, each student was given two coins and asked to find the most likely outcome when the two coins were flipped. After the two-coin activity a three-coin problem was discussed. Although the situation for item K8 was changed from flipping coins to pulling out marbles, some students recognised the parallel structure between the two problems. Student j8081141 correctly chose option (c) for J8 with the reason that *for each spinner, the odds for the two coloured parts was 50% to 50%*. When he was asked to explain the reason further, he said he could not say any more. For item K8, he indicated that *there're four possible outcomes of the experiment…. The probability for one marble is white and the other one is black is ½, bigger than the probability for two white marbles or two black marbles, which each has 25% probability.* When he was asked how he got the chance values, he said *from the activity of flipping coins.*

Most of the students who were able to recognise the parallel structure between the activity and item K8 were also able to answer J8 correctly after the teaching intervention. However, a few students were found who correctly answered K8 but when they were required to answer J8 again they could not identify the parallel structure or needed a hint. For example, before the teaching intervention, student j8082130 chose (d) (equiprobable option) for J8 because she thought the final position of the arrowhead *depends on the force* *of the spin*. In her written answer for item K8 she used equiprobability again and assigned a 1/3 chance to each of the three events but then she crossed it out and changed to a correct response (E level response) based on correct chance values. In the interview, before being shown her written reply for K8, she was asked to do item J8 again and she selected option (d) again using the equiprobability (G5). Then she was shown her written answer for K8 and was asked why she crossed out her original reason, she said it was because she remembered the activity and its result. When she was asked whether she thought J8 and K8 looked alike, she said *"aha" and the answer for J8 should also be (c)*.

The ability of some students to identify the parallel structure between the two items provides evidence to support the hypothesis that with the appropriate approach students are able to transfer information from one probability situation to another.

<u>*Change in reasoning*</u>

Student j8081080 correctly chose option (c) in the two tests but the reasons were quite different, one used a misconception and other was based on her correct intuition. For item J8, she explained her answer by saying that it was *because the strength of the two hands is not stable.* She thought the two arrowheads would stop in different places because of the variation of strength in the hands. It was coded as using a subjective judgement misconception (G1). For item K8, she explained that it was *because the marbles in the two bags have the same ratio, it is most likely that one marble is white and the other one is black.* In the interview, when she was asked to elaborate her explanation, she said *it's impossible to easily pull out two white* *marbles or two black marbles. One white and one black is more likely to happen and is also fair to each colour.*

As with the outcome approach, it was also found that the use of equiprobability (G5) was difficult to be overcome through a short-term teaching intervention. The difficulty of overcoming this misconception is consistent with the literature (for example, Lecoutre, 1992). In this study, 10 out of 19 students who used the equiprobability for J8 did not use it for K8. But another four students who used misconception G4 (chance cannot be measured mathematically) for J8 changed to G5 for K8.

Student j8081090 selected the equiprobable option (d) for both items J8 and K8. For J8, she thought the two arrowheads would stop randomly, so the likelihood of each outcome should be equiprobable. For K8, she indicated further that the likelihood of each outcome should be 1/3. In the interview, when she was asked whether she still remembers the activity of flipping coins, she said she still remembered it. *In the activity, it was found that the most likely outcome was one head and one tail. The probabilities we found were 1/4, 1/4 and 2/4.* Although she said she felt the two problems look like each other, she did not want to change her answer for item K8.

One point worth noting is that in the pre-test some students felt confused about the difference between the last two options in the item, (d) and (e), as they thought they were the same. Furthermore, some students chose the equiprobable option (d) but they did not know the common value. This suggests a hypothesis that for some students, when they say two outcomes are equally likely to happen, it does not mean they really think the chance of the two outcomes can be measured mathematically by an equal number. Here are two examples to support this hypothesis.

Student j8081201 chose option (d) for item J8 with the reason *since all of them are possible.* In the interview, he said actually that options (d) and (e) were almost the same. The interviewer explained that option (d) meant each event had a 1/3 chance and option (e) meant no chance values could be assigned to the three events. In light of this new information he said he would now chose option (e).

Student j8082050 chose option (d) for item J8 with a similar reason to that has been mentioned above. In the interview, she was asked since the three outcomes had the same likelihood, whether she knew the value of the likelihood. She said she did not know the value. In order to investigate how she got the equiprobable conclusion she was given a new question where both the spinners had more blue sectors like item IV8 used in the main study. She changed her equiprobable answer to another option (it is most likely that both the arrowheads stop in the blue part) using the compound approach (G11). So this reason was not coded as G5 (equiprobability) but as G11. She selected the equiprobability answer for J8 just because she believed the combinations of the equally likely outcome at each stage have an equal chance (see chapter 4 Group 11).

These two examples suggest again that further investigation of the equiprobability misconception is needed.

Overview

In order to have an overall picture of the change resulting from the teaching intervention, the above information is classified and summarised again according to class. As was done with the data in Chapter 4, the following report on the students' misconceptions is descriptive. For each misconception, the students were divided into two groups: those who never used the misconception and those who used the misconception at least once. Then the three most common misconceptions were listed. The result of this organisation is shown in Table 6.5.

Table 6.5 The three most common misconceptions observed in pre-test and post-test

<table>
<tr><th colspan="2">Class</th><th>First main misconception</th><th>Second main misconception</th><th>Third main misconception</th></tr>
<tr><td rowspan="2">Pre-test</td><td>W/ 26</td><td>Group 5: (69%*)
Equiprobability</td><td>Group 4: (62%)
Chance cannot be measured</td><td>Group 6: (54%)
Outcome approach</td></tr>
<tr><td>W/O 23</td><td>Group 4: (57%)
Chance cannot be measured</td><td>Group 5: (52%)
Equiprobability</td><td>Group 6: (52%)
Outcome approach</td></tr>
<tr><td rowspan="2">Post-test</td><td>W/ 26</td><td>Group 5: (46%)
Equiprobability</td><td>Group 11: (35%)
Own methods in chance comparison</td><td>Group 6: (31%)
Outcome approach</td></tr>
<tr><td>W/O 23</td><td>Group 6: (43%)
Outcome approach</td><td>Group5: (39%)
Equiprobability</td><td>Group 14: (30%)
Own methods in chance calculation</td></tr>
</table>

* All the percentages in this table indicate the percentage of students who used each misconception at least once

Comparing Table 6.5 with Table 4.27 (in chapter 4) it was found that the three most common misconceptions observed in the main study were the same as that observed before the teaching intervention. However, the percentages and order were different. This was mainly because, as was explained at the beginning of chapter 4 and before Table 4.27, the data reported in that table was an underestimate of the actual number of the students with the misconceptions. In addition, Table 4.27 summarised the data collected from 209 students in two school streams, but Table 6.5 only contains the data from 49 ordinary stream students.

Table 6.5 shows that before the teaching intervention between 50% and 70% of the students used the misconceptions G4, G5 and G6 at least once. The class without a computer used the three misconceptions slightly less often than the class with the computer. After the teaching intervention, between 30% and 50% of the students used the misconceptions G5, G6, G11 and G14 at least once. The class without a computer still used G4, G5 and G11 slightly less often than the class with the computer but used G6 and G14 a bit more often than the other class. Overall this means that although the teaching intervention helped many students overcome the misconception of chance cannot be measured mathematically (G4), there were still many students with the misconception of equiprobability (G5) or the outcome approach (G6). The misconception G11 (used own methods in chance comparison) was used slightly less often in both classes after the teaching intervention. However, since the use of the misconception G4 decreased considerably, G11 became one of the main misconceptions observed in the post-test. After the teaching intervention, students from both classes made a greater effort to calculate, especially in the class without a computer. Because of the students' limited knowledge of how to calculate theoretical probability, their own methods of chance calculation often conflict with probability theory and these responses were coded as using the misconception G14.

Compared to the data obtained before and after the teaching intervention, it was found that all the misconceptions, except G14, were used less often by the students after the teaching intervention, especially misconception G4. In the class with the computer, the greatest improvement in overcoming misconceptions was for G2 (example-based interpretations for possible and impossible) and G4. The percentage of students who used G2 and G4 at least once decreased from 35% to 0% and from 62% to 19% respectively. In the class without a computer, the greatest improvement was for G1 (subjective judgements) and G4. The percentages using them at least once decreased from 30% to 0% and from 57% to 4% respectively. While, in the two classes, after the teaching intervention, the percentage of students who used G14 at least once increased from 4% to 23% (W/) and 30% (W/O).

The analysis on the students' developmental structure includes two approaches, descriptive and numerical, which parallels the approach in chapter 5. Table 6.6 indicates how often each level's responses were observed in each class. After that, using the programme RUMM, each student was assigned a numerical understanding location. The distinction between the time interval and between the two classes was examined.

Table 6.6 Overall picture of students' SOLO levels in pre-test and post-test

Percentages of responses in each level	Pre-test		Post-test	
	W/ 26	W/O 23	W/ 26	W/O 23
Prestructural	0%	2%	0%	0%
Unistructural	33%	31%	3%	2%
Multistructural	58%	53%	55%	49%
Relational	8%	14%	29%	30%
Extended abstract	1%	0%	13%	19%

Table 6.6 summarises how often each SOLO level response was observed. Comparing it with Table 5.6 in chapter 5, the data before the

teaching intervention were quite similar with the data for the ordinary grade 8 students reported in the main study. About 50% of the responses before and after the teaching intervention were M responses. The main change was the decrease of U responses and the increase of R and E responses.

The second approach of analysing the change in students' understanding was comparing the two locations before and after the teaching intervention. The result is shown in Table 6.7.

Table 6.7 Overall picture of students' numerical understanding location index* before and after the teaching intervention

Number of students located in each interval	Pre-test		Post-test	
	W/ 26	W/O 23	W/ 26	W/O 23
(-3, -1.2]	1 (4%)	1 (4%)	0 (0%)	0 (0%)
(-1.2, 0.6]	14 (54%)	12 (53%)	1 (4%)	1 (4%)
(0.6, 2.4]	11 (42%)	10 (43%)	12 (46%)	6 (26%)
(2.4, 4.2]			10 (38%)	12 (53%)
(4.2, 6)			3 (12%)	4 (17%)
Mean and Standard Deviation	Mean=0.271 SD=0.942		Mean=2.827 SD=1.317	

* The higher the location the better the students' understanding

From Table 6.7 and 6.8, the improvement of the students' locations (location gains) between the two tests is clear. After the teaching intervention the mean for all students' location value was higher but the variance among the students was larger. A t-test also shows that the gain is statistically significant (t = 13.98, p = 0.000). Between the two classes, no statistically significant difference was observed before the teaching (t = -0.06, p = 0.95) or after the teaching (t = -1.29, p = 0.20). As for location gains, after the teaching, there are many more students in the class without a computer located above the mean. However, a t-test shows that there was no significant difference between the two classes in the location gains from pre-test to post-test at the 0.05 level (t= 1.27, p = 0.21).

Table 6.8 Mean and SD of the locations, location gains in the two classes involved in teaching intervention

Classes	Pre-location		Post-location		Location gains	
	Mean	SD	Mean	SD	Mean	SD
W/ 26	0.26	0.93	2.60	1.40	2.34	1.34
W/O 23	0.28	0.98	3.08	1.20	2.80	1.19

This result was expected since the teaching conditions in the two classes were the same except for the way the data for a long series of experiments was generated. The class with the computer was exposed to computer simulations, while the other class received the information from the given data printed in the workbook. In practice, the two classes did the same things most of the time.

Further Information

It was found that context played a role in some students' answers. The student j8082201 who gave the only one R response for J1 illustrates this point. In the post-test, he based his reasons for K6, K7 and K8 (slip or marble items) on chance calculation but applied equiprobability in answering K9 (spinner item). When he was asked why he calculated in other chance comparison items but did not do any calculations in K9, he said when an item involved numbers he thought he should calculate, otherwise, the calculation was unnecessary. However, most of the other students gave a higher response for K9 than for K6, K7 and K8, since the ratio between area of the two colour parts was visualised in K9 and it made proportional comparison much easier.

As was mentioned in chapter 4 this research has separated equiprobability responses into each outcome with 50% chance, all outcomes equally share an equal chance and close chances are the same in practice (they are called version I, II and III in the following explanation). Such a distinction has rarely been mentioned in the literature. In the teaching intervention it was found that 15 out of 49 students only used version I, either in one or both tests. The data for version II and III were seven out of 49 and one out of 49 respectively. Only three out of 49 students used more than one version in one test. Another ten students changed from one version to another, such as changing from version I to II after the teaching intervention. This result and the explanation given at the end of category III indicated that equiprobable answers might be the result of different reasoning, so further research on equiprobable responses is necessary in order to inform appropriate remediation.

In the teaching intervention, the ideas of chance might be measured mathematically and observed frequencies obtained after a long run can be used as an estimation of probability were emphasised. The classical definition of probability and its relation to the frequentist definition was not written down in the workbook, but was discussed with the whole class after each activity. When looking at the final estimated value for the chance of target event, the students were asked to think about it and explain why it should be that value. Only some of the students (often they came from the class without a computer) could initially give an explanation, but during the discussion, more students understood the reason. It was found that in answering K7 (where proportional reasoning was required) 36 of the 49 students were able to use proportional reasoning. Among the 36 students, 15 were able to answer using classical probability values, three (all in the class with the computer) said they did not know the values, five gave wrong values (such as taking part-part ratio instead of part-whole ratio as the measure of probability) and the remaining 13 students were not specifically asked about this in the interviews. This finding supports the hypothesis suggested by Fischbein and Gazit (1984) namely, that

> Probabilistic thinking and proportional reasoning are based on two distinct mental schemata. A progress obtained, as an effect of instruction, in one direction does not imply an improvement on the other. (p. 23)

The eight students mentioned above could use proportional reasoning in chance comparison but actually they did not know the real chance values.

Summary

Supported by the data shown in this chapter, one may conclude that this activity-based short-term teaching intervention (six lessons) given to a small class (about 25 students) helps these ordinary school grade 8 students understand probability better and overcome some of their misconceptions about probability. After the teaching intervention it was found that the students in the two classes had an improvement in their answers and reasoning but no statistically significant difference was found between the classes. Misconceptions, such as, chance cannot be measured mathematically, subjective judgements and example-based interpretations for possible and impossible, had virtually disappeared or were used much less after the teaching intervention. However, as measured by the items in this study it would appear that the outcome approach and equiprobability were relatively stable and not easily eliminated.

The data obtained in both the main study and the teaching intervention suggests that the study of one approach of probability (theoretical approach or empirical approach) does not imply an improvement on the other. In the main study, grade 12 students who had been taught in the theoretical approach did not do very well in items relating to empirical probability. In the teaching intervention, the students were taught using the empirical approach but their knowledge on theoretical probability was weak. Therefore we cannot use one approach to replace another approach. After providing students with some experience of uncertainty and using observed frequencies to estimate probability, the theoretical definition of probability and using proportional reasoning in chance comparison needs to be emphasised further.

Chapter 7 Summary, Implications and Recommendations

The final chapter of this thesis includes three components. The first is a summary of the study. Its methods, important findings and conclusions are briefly reviewed. The second section indicates some didactical implications of the findings. At the end of this chapter, associated with both the limitations and findings of this study, some further topics for research are recommended.

Summary of This Study

The major focus of this study was on Chinese students' understanding of the concept of probability, both the classical and frequentist definitions. Specifically, there were three main research questions:

1. What are the main misconceptions of Chinese students when answering chance interpretation and chance comparison problems?
2. What is the developmental structure of students' understanding of probability?
3. Can an activity-based short-term teaching programme improve Grade 8 students' understanding of probability?

Eighty-three items, distributed among 9 questionnaires, plus selected interviews, were used in answering the first two research questions. The items covered four general categories: Category I - Identification of impossible, possible and certain events; Category II - Interpretation of chance values; Category III - Chance comparison in one-stage experiments; and Category IV - Chance comparison in two-stage experiments. The third question was answered by a short-term teaching intervention, together with pre-test and post-test questionnaires and interviews of all the students. The summary is divided into three sections: the pilot study, the main study and the teaching intervention.

Pilot Study

In this stage of the study, the items used for data collection were developed. The development of the items was based on the literature. In each category two to three items used by other researchers were selected as the basis for development. They were adapted in a variety of ways in order to make them appropriate for this study.

Context. The context for some items was changed. For example, Chinese students were unfamiliar with the idea of a tombola game so a parallel situation involving playing cards was used.

Format. Some of the questions in the literature used an open-ended format. It was felt that some students might have trouble giving enough information on their thinking when asked to provide written responses to open-ended questions. So they were adapted to a multiple-choice format. Also, specific options were developed to allow the further investigation of misconceptions such as the equiprobability bias.

Parallel Items. This study included an investigation of the role that context and data played in students using misconceptions. In order to do this parallel versions of items with the context and data changed were developed. For example, to investigate the role of context, in addition to the items of picking out names from a box, items of picking out marbles from a bag and spinning arrowheads of spinners were also used in parallel items for chance comparison. For the investigation of data, for example, the difference between the number of boys' and girls' names was changed from equal to close to far apart.

After the initial development of the items and prior to any piloting they were sent to overseas experts for comment. Their comments were incorporated into a revised set of items to be piloted. Finally, all the items were translated into Chinese for administering the questionnaires. The Chinese version was given to another two experts in Singapore to check the accuracy of the translation.

The pilot sample was 144 students in grades 6, 8 and 12 in a small city near Shanghai. The 48 subjects in grade 12 came from an advanced high school and the other 96 subjects in grades 6 and 8 came from two ordinary high schools. The test papers were sent to a colleague in China, together with specific instructions on how to administer the tests. These instructions asked her to show students a six-sided die; show them a pack of cards and remove all the picture cards; tell them that if they did not agree with any of the multiple-choice options they could give an alternative; and not provide any help to the students if they asked questions.

The results of the pilot study showed the viability of the test both in terms of the time frame and structure. Based on the results of pilot study together with the original suggestions by the experts some final revisions were made to clarify the items. A total of 83 items were prepared for the main study (see Appendix A), organised into nine disjoint questionnaires for administration. Most of the items in the main study were in a multiple-choice format and students were asked to give reasons for all their responses. In addition, they were asked to indicate their level of confidence in their answers.

Main Study

The first two questions were investigated by the main study. The items were organised into nine distinct questionnaires which were administered in 1998 to 567 high school students in Shanghai over several weeks. All the questionnaires were used in all the classes with each student answering one questionnaire. The students came from grades 6, 8 and 12 in different school streams, advanced or ordinary. The grade 6 and 8 students had no prior experience with studying probability in school, but the students in grade 12 had studied a brief introduction to probability of about 8 hours. The approach used in the unit was mainly theoretical probability and rarely involved empirical activities.

A subset of 64 out of the 567 students was interviewed. Each student was interviewed on the day after he or she answered the questionnaire. The selection was made by the researcher quickly reviewing the written responses, with the main basis for choosing students to interview being that the reasons they gave were unclear. The purpose of the interviews was to clarify students' thinking as indicated in their written answers. The initial organisation of the interview was to give the students their answer script and ask them to answer some selected questions again and explain what they had written. Where appropriate, students were also asked parallel questions to clarify their thinking and determine if the thinking was consistent. Some other students who had given interesting or unusual responses or who had not given written reason were also interviewed. All the interviews were audiotaped.

Main findings for the first research question

There are three aspects to the findings associated with this research question. The first concerns the existence of misconceptions, the second concerns the role that grade and stream play in the use of misconceptions and the third concerns the role that data and context play in students use of misconceptions.

First, many misconceptions were observed in this study and they were combined into fourteen groups: (1) subjective judgements; (2) example-based interpretations for possible and impossible; (3) possible means certain; (4) chance cannot be measured mathematically; (5) equiprobability; (6) outcome approach; (7) one trial is unrelated to other trials; (8) interpreting chance by data matching or word matching; (9) increasing repetitions is not better for predicting; (10) positive and negative recency; (11) used own methods in chance comparison; (12) taking different order as the same; (13) misuse or extend conclusions inappropriately; and (14) used own methods of chance calculation. The outcome approach, chance cannot be measured mathematically, the compound approach and equiprobability are the four main misconceptions observed in each grade and each stream. In this study, only those responses that clearly indicated that a misconception was used were counted. Even with this conservative criterion the first two main misconceptions observed in each grade, were used by more than 30% of the students. Furthermore, all the groups of the misconceptions were observed in each grade and each stream.

Comparatively, more misconceptions associated with the frequentist definition of probability were observed. This might be because the students had had very limited experience, if any, of experimental probability since the curriculum rarely included probability experiments. The literature also indicates that frequentist probability is a weak area of students' understanding (for example, Green, 1983a). Specific results for the four most common misconceptions are described now.

Outcome approach. 43% of grade 6 students, 39% of grade 8 students and 31% of grade 12 students clearly used the outcome approach at least once in their written responses. In addition to students using 50% as a guide to determine the certainty of an event happening (Konold, 1989), many students in this study used a parallel or slightly more sophisticated criterion as a guide, that is, comparing the chance of target outcome happening and that of it not happening. For example, when they were informed that the weather forecaster predicted that there was an 80% chance of rain, they concluded that there was a 20% chance that it would not rain. Then they compared the chances for the two outcomes and concluded that it would rain. Further, an extension of the outcome approach, named as the weak outcome approach in this study, seems to exist. This misconception leans towards the outcome approach, but with a small adjustment. For example, when students were asked to interpret the meaning of an 80% chance of rain, students using the weak outcome approach preferred to choose 9 out of 10 such days it would rain, rather than choose it would certainly rain that day or it would rain 8 out of 10 such days. They are biased towards the outcome approach but do not have the level of certainty associated with the outcome approach, per se.

Chance cannot be measured mathematically. Forty-six percent of grade 6 students, 35% of grade 8 students and 19% of grade 12 students clearly used the misconception at least once in their written responses. There seems to be several justifications that lead students to this misconception. The first belief is that chance is the same as luck. The second is that chance changes with every trial. The third is that assigning a greater chance to an event just means to predict it **will** happen. Each of these justifications led students to believe that you cannot measure the chance of a possible event. When students used this misconception they usually concluded that it was impossible to compare the likelihood of all the possible outcomes. Since the second and the third justifications bind the measurement of chance to a particular trial and students concluded that the event with the greater chance was more likely to happen, it seems that this misconception has some commonality with the outcome approach.

It should be noted that some grade 12 students who were able to calculate the probability of an outcome occurring still used this misconception. Some of them believed that for one specific trial, such as picking a name out of a box, probability is meaningless. Probability is only meaningful for a series of trials, so for a specific trial, it is impossible to compare the likelihood of all the possible outcomes. Some of them did not believe in probability, as they found that the results from probability theory were in conflict with the real data. For these students their real-world experiences overrode the logical conclusion that should be drawn from the calculation.

Compound approach. In solving multi-stage chance comparison problems many students tended to split a multi-stage experiment into several distinct one-stage experiments then compound them intuitively. This approach is of particular interest in this study and it is referred to as the compound approach. For example, consider the two-stage experiment of drawing one marble from each of the two bags, each of which contains some black marbles and white marbles. If in each bag there are more black marbles than white marbles students using this misconception believe that since drawing a black marble is more likely for each bag then drawing out two black marbles is most likely for the two-stage experiment. They compound the results for each stage without any calculations. This approach has been reported but rarely been studied systematically. Lecoutre and Durand (1988) and Fischbein et al. (1991) used the same item involving rolling two ordinary six-sided dice. They observed the compound approach when the outcomes in each stage were equally likely to happen. Konold et al. (1993) used both items where the outcomes in each stage were equally likely to happen (flipping a fair coins five times), and where they were not equally likely to happen (rolling a painted die six times which had 5 sides black and 1 side white). In this study, like Konold et al. (1993), both situations (equally likely and unequally likely) were included, but with two stages and the probabilities in the two stages were varied. For example, in the marbles items, equal numbers of black and white marbles were put in a bag, or different amount but with the same ratio of marbles or different amount with different ratio of marbles were put in. Parallel items involving spinning the arrowheads of two spinners were also used.

Twenty-seven percent grade 6 students, 17% grade 8 students and 34% grade 12 students clearly used the compound approach at least once in their written responses. When comparing the numbers of items that were designed primarily to investigate chance cannot be measured or equiprobability (30 items) or the outcome approach (27 items), far fewer items (8 items) were designed to investigate this misconception. However, it was still observed very frequently in this study. This suggests that this particular misconception is probably worthy of further research.

Equiprobability. About 20% of the students in each grade clearly used the equiprobability at least once in their written responses. In the literature there has been an extensive discussion of equiprobability (for example, Lecoutre, 1992). The usual situation in the literature involves two outcomes to be compared, such as pulling out a boy's or girl's name. The items used in this study allowed the researcher to extend the investigation of this misconception to a situation where students had to compare the likelihood of more than two outcomes. The data show that students used different versions of equiprobability. The first was the situation where each outcome is assumed to have a 50% chance. For example, when drawing a name from a box with 5 boys and 27 girls names some students believe it is equally likely to get a boy's or a girl's name, both have a 50% chance. The second involved a situation where there were n outcomes and the students concluded that all n outcomes equally share an equal chance, 1/n. For example, when tossing two coins some students believe the three events, a head and a tail, two heads or two tails are equiprobable, with the chance of each occurring being 1/3. The third involved situations where the numbers were close and students concluded that they were the same in practice. For example, picking a name out from a box with 22 boys' names and 20 girls' names. Some students believe it is equally likely to pick out a boy’s name or a girl’s name arguing that the number of boys and girls is close so, in practice, the likelihood is the same. In addition to these three reasons, the compound approach can lead to equiprobable responses when the outcomes in each stage are equiprobable, as has been illustrated previously. However, in this study such equiprobable answers were coded as using the compound approach and not equiprobability.

The data of this study show that compared to other research the use of equiprobability was not as common for Chinese students as for students in the West. There are a variety of possible explanations. One might be due to the conservative approach of coding. In order for a response to be coded as using equiprobability both the answer and reason had to clearly reflect this reasoning. This means that incomplete or unclear answers were not coded as using the misconception. Another reason could be the use of alternative options in the questions as well as coding different misconceptions. For example, the option “it is impossible to compare the likelihood of the two outcomes” was added and allowed for the identification of the misconception of chance cannot be measured mathematically. More than 30% of grade 6 and 8 students in this study selected this option. Together with findings that the use of this misconception is data and context dependent means that variability between research studies might be expected.

The second main finding associated with the first research question was that except for six groups of misconception, where there were only a relatively few students who used misconceptions in these groups, misconceptions were used less frequently by grade 12 students than by grade

6 and 8 students. Also the misconceptions were generally used less frequently or with equal frequency by advanced school students when compared to ordinary school students. The data for the four main misconceptions (see Table 7.1) is now investigated in more detail.

Table 7.1 Percentage of students who used the four main misconceptions at least once, classified by grades and streams

The four main misconceptions	Grades			Streams	
	Gr 6	Gr 8	Gr 12	Ord	Adv
Outcome approach	43%	39%	31%	38%	37%
Chance cannot be measured	46%	35%	19%	36%	29%
Equiprobability	20%	19%	18%	18%	20%
Compound approach	27%	17%	34%	24%	28%

It seems that except for the compound approach, the other three main misconceptions decrease with age. The fall is very obvious for chance cannot be measured mathematically, but the decline is marginal for equiprobability. This decline indicates that the misconception of chance cannot be measured mathematically can be overcome easily but the outcome approach and equiprobability are more stable. The compound approach occurred less often in grade 8 than in grade 6, but occurred much more often in grade 12 than in either grade 6 or 8. The increase in grade 12 might be because fewer grade 12 students thought chance cannot be measured mathematically and instead they thought that the compound approach worked in solving two-stage experiment items.

For the other groups of misconceptions, it was also found that subjective judgements and example-based interpretations for possible and impossible decreased with age. The two misconceptions belong to the lowest levels of understanding of probability. For most students, they can correct these misconceptions by themselves when they have more experience in dealing with uncertain situations. The misconceptions of interpreting chance by data matching or word matching and using own methods in one-stage chance comparison were used much less often by grade 12 students than grade 6 or 8 students. The misconception of increasing repetition is not better for predicting was used less often by grade 8 and 12 students than grade 6 students.

Examining the role of stream, it was found that if all three grades were combined, almost no difference (no more than 2%) was observed between ordinary students and advanced students in using the following four groups of misconceptions at least once: subjective judgements, equiprobability, the outcome approach and increasing repetition is not better for predicting. More ordinary students used example-based interpretations for possible and impossible (14% vs. 5%), chance cannot be measured mathematically (36% vs. 29%) and interpreting chance by data matching or word matching (15% vs. 11%). Only the compound approach was observed more in advanced students than ordinary students (28% vs. 24%). The advanced students, who should be better academically, might well try and generate a reasonable method of solving the problem, resulting in more compound approach misconceptions. On the surface the compound approach seems intuitively reasonable.

In all, except for six groups where an analysis of the role of stream was not reasonable, most of the other eight groups of misconceptions were used less often or equally often by older students and advanced school students.

Only the compound approach was used more often by grade 12 students and advanced students.

The third main finding for the first research question was that context and data played a role in eliciting some misconceptions. For the four main misconceptions, the use of chance cannot be measured mathematically does not seem to depend on data or context. A Rasch analysis indicated that when the context of the items was changed from picking out a marble to spinning an arrowhead of a spinner, or when the total names in a box was changed from a smaller number to a larger number or decrease/increase in the difference of boys' and girls' names in the box, no major shift in the items' locations was observed. Since the students believed chance cannot be measured, no matter what context involved or what data given in the items, they used it consistently.

The compound approach is influenced by data but not influenced by context. The approach was used more often when the compositions of bags/spinners were quite different from each other than they are equal or close. Actually, when the compositions were quite different from each other, the results from using the compound approach are in agreement with those that could be deduced from probability calculations.

The outcome approach is influenced by context but may not be influenced by data (whether using a 50% chance elicits more outcome approach responses is still unclear). Less outcome approach responses were observed in a context involving marbles than ones involving football or the weather. It seems in situations where an event is affected by factors out of a person's control it is easier to interpret probability as an indication of whether an outcome will happen in a single trial rather than as an indication of how often it will happen when the event is repeated.

Both context and data affect the equiprobability bias. More equiprobable responses were observed in drawing names' items than drawing marbles or spinning arrowhead items, and also more equiprobable responses were observed when the compositions were close than when they were far apart. It may be that students' familiarity with drawing a name, where everyone has an equal chance (all names are equiprobable) leads to more equiprobability misconceptions. When the data are close some students use the specific version of equiprobability, namely that if the chances are close they are the same in practice, resulting in more equiprobability responses.

The misconceptions of physical properties based judgements (subjective judgements) and one trial is unrelated to other trials are influenced by context but not by data; egotistical beliefs based judgements (subjective judgements), example-based interpretations for possible and impossible and possible means certain are influenced neither by context or data. However, since these misconceptions were not observed as often as the four main misconceptions, these conclusions should be considered tentative. For the other groups of misconceptions, such as, misuse or extend conclusions inappropriately, as even fewer numbers of students used them, the role of data and context was not investigated.

Overall, it appears that the role of data and context is complex. For some misconceptions both impact on the use of a misconception, for others only one plays a role and in some cases neither impact on the use of a misconception. This is an area that is worth considering for expanded and systematic research.

Main findings on the second research question

All the responses, except for about 4% that could not be understood or where students answered another question that could not be coded, were labelled with a SOLO level. The five levels of concrete operational mode are: prestructural (P), unistructural (U), multistructural (M), relational (R) and extended abstract (E). Blank, fully irrelevant, illogical, egotistic answers or inability to become engaged in item answers were coded as P level responses. It was found that the SOLO taxonomy was a useful model for describing students' responses, enabling the identification of hierarchical understanding levels for the concept probability. In this study, each student was assigned two understanding indices. One was a descriptive label, formed by a set of SOLO codes for all his or her written responses. The other was a numerical label, assigned by a Rasch analysis based on the descriptive index. The analysis of the data shows there are three main findings associated with the second research question.

First, in general, there is no improvement in the developmental level between grades 6 and 8, the two grades without any formal probability training. Grade 12 students have a better understanding than the younger students. Table 7.2 shows the percentages of R or E level responses observed in different grades and streams.

Investigating the data by streams it can be concluded that ordinary school grade 6 and 8 students had the poorest understanding of probability, advanced school grade 6 and 8 students had a slightly better understanding, but not as good as ordinary school grade 12 students. Advanced school grade 12 students had the best understanding of probability among all the students.

Table 7.2 Percentages of R or E level responses observed in different grades and streams

Streams	Grade 6		Grade 8		Grade 12	
Ordinary	7%	8%	8%	11%	19%	27%
Advanced	13%		15%		35%	

These conclusions might be expected, since grade 6 and 8 students had no instruction in probability and grade 12 students were mature in age but had also been taught probability. Advanced school students were selected by a streaming examination after grade 5 and after grade 9. Their academic background was usually much better than the ordinary school students.

Second, the variation between the students with and without formal training in probability appears to be less pronounced for category II items (interpretation of chance values) and more pronounced in category IV items (chance comparison in two-stage experiments). This was mainly because the training was focused on probability calculation and it enabled a large number of grade 12 students to answer category IV items at the R or E level. However, all the students in this study had very limited experience with probabilistic experiments, so their response levels were generally lower when they were required to interpret chance in a frequentist approach.

Third, even when students can give correct responses their reasons are often not very sophisticated. For grades 6, 8 and 12, the percentages of responses at the R or E level were 8%, 11% and 27%, respectively. In order for a response to be at the R or E level, students need to use part-part ratio or part-whole ratio in chance measurement, use proportional reasoning in chance comparison, know a larger number of repetitions is a more reliable predictor or express the idea of making a few or a lot of repetitions to find trends across sampling. One possible reason for the comparatively few R and E level responses is that many items can be answered without calculating probabilities and the calculation was not required in the items.

Data coding in the main study was only based on students' written responses. The information obtained in the interviews was reported but not combined in data coding.

Teaching Intervention

The third research question was answered by the teaching intervention. The short-term teaching intervention was done in an ordinary school in Shanghai in 1999. Two grade 8 classes, one with 25 students and the other with 26 students, were used for the teaching intervention. The students' ages were 13-14 years. They had no prior experiences in studying probability before they were involved in this study.

The instruments used in the pre- and post-test were parallel. Most of the items were selected directly from the 83 items developed for main study. Multiple-choice plus explanation was still the main format for the items but there was only one pre-test and one post-test. All the students were interviewed before and after the two tests and all the interviews were audiotaped.

The teaching intervention involved six activity lessons (40 minutes per lesson) to each class, twice a week. The teaching was activity-based combined with whole class discussion. The activities were designed based on the information obtained from the main study. The intervention focused on the misconceptions related to identification of impossible, possible and certain events and the frequentist definition of probability. One class had access to a computer for simulations while the other did not. However, most of the conditions such as the activities, workbook, the problems for whole class discussion and the teacher were the same for the two classes. The main difference was that after students completed activities with dice, coins, and so on, in the computer lab class the data for a long series of experiments was simulated in front of the students, while in the other class the students were given the data for a long series of experiments and were told that they were computer generated. There are four major findings associated with the third research question reported in this summary.

First, an activity-based short-term teaching programme (six lessons) with a small class (about 25 students) can help ordinary school grade 8 students improve their performance in terms of giving correct answers. After the teaching, the percentage of correct answers increased from about 50% to about 80%.

Second, there was a reduction in terms of the misconceptions used after the teaching intervention. All the fourteen groups of misconceptions except groups 13 (misused or extended conclusions inappropriately) reported in the main study were also observed in the teaching experiment. Equiprobability, chance cannot be measured mathematically, the outcome approach and own methods in chance comparison were the most common misconceptions prior to the teaching intervention. Equiprobability, outcome approach, own methods in chance comparison and own methods in chance calculation were the most common misconceptions after the teaching intervention. Although, equiprobability, outcome approach and own methods of chance comparison were still the most common, they all occurred less often after the teaching intervention. The percentage of students who used them decreased from 61% to 43%, from 53% to 37% and from 35% to 31%, respectively. The misconceptions of chance cannot be measured mathematically, subjective judgements and example-based interpretations for possible and impossible were easily overcome. After the teaching intervention, the percentage of students who used them decreased from 59% to 12%, from 14% to 2% and from 31% to 4%, respectively. As the data show although the occurrence of misconceptions of the outcome approach and equiprobability decreased, it is not easy to eliminate them by the use of a short teaching intervention. This resistance of these misconceptions to change has been reported in the literature (for example, Lecoutre, 1992, Konold 1989).

There was one misconception (own method of chance calculation) that was used more often in both classes after the teaching. The percentages increased from 4% to 23% (class with the computer) and from 4% to 30% (class without the computer). Since the teaching focused on the frequentist not the classical definition of probability, students' knowledge of how to calculate theoretical probability was limited. Their own methods of chance calculation were often in conflict with probability theory.

Third, there was an increase in terms of the developmental level of thinking. According to students' numerical understanding index, after the teaching, their location gain is statistically significant ($t = 13.98$, $p = 0.000$).

There was no significant difference between the two classes in the location gains from pre-test to post-test (t = 1.27, $p = 0.21$). It would appear that even a short teaching intervention can improve a student's understanding.

Fourth, students who used proportional reasoning in chance comparison might not be aware of how to measure probability. During the interview, eight students were found who used proportional reasoning in chance comparison but actually they did not know the real chance values. It appears that studying frequentist probability does not necessarily contribute to students' knowledge of classical probability.

In all, based on the main study and the teaching intervention, the following conclusions can be drawn: (1) Outcome approach, chance cannot be measured mathematically and compound approach are three main misconceptions of probability in Chinese students independent of school streams or background in probability; (2) Students' understanding of probability does not improve naturally with age - teaching plays an important role; (3) A short teaching intervention can help ordinary grade 8 students overcome some misconceptions and improve their understanding of frequentist definition of probability; and (4) Context and data involved in an item does play a part in eliciting some misconceptions.

The focus of the next section is on the didactical implications of the research.

Didactical Implications

The question of the implications is important for any research study. Although this section is brief, in the opinion of the researcher it contains important implications for the teaching of probability, particularly in Chinese schools. The implications have been organised under two headings, implications for curriculum developers and implications for teachers.

For Curriculum Developers

In China, research on students' misconceptions has not been emphasised and its role in curriculum development, teaching and learning has not been widely applied. Development of formal curriculum is mainly based on the current syllabus, experiences obtained from past reform and the current formal component of the curriculum in other countries. How to teach a topic and the main misconceptions of students are mentioned in teacher guides, but usually they are a summary based on teachers' experiences, and do not reflect the available global research information.

However, current practise is under review. Some new topics that have rarely been taught at the school level are being introduced or will be introduced in near future. These include topics such as probability, statistics and calculus. Some topics, such as negative numbers are in the process of being introduced earlier in the curriculum. The nature of these changes means that for many topics there is no prior experience within the Chinese context, either in terms of the topic being part of the school curriculum or in terms of the topic being taught to students at a given age. In making the changes, it is important to be aware of the available research information. Based on this study, the following three suggestions for the topic of probability are proposed for consideration by Chinese curriculum developers.

First, it is possible to introduce probability effectively at an earlier age by an activity-based teaching. At the present time, the teaching of probability is virtually non-existent within the national curriculum of China. Where it is taught it is at the end of high school and the approach is heavily biased to the classical definition of probability. However, according to the results of the teaching experiment undertaken in this study and the literature (for example, Fischbein and Gazit, 1984), it is practical to introduce both the frequentist and classical definitions of probability in junior high school.

Second, to concentrate all the probability teaching in one programme in one semester does not help overcoming some stable misconceptions that students have developed from a relatively early age. This study and the literature show that some of the misconceptions exist before formal teaching, some new misconceptions might appear during teaching and some of the existing misconceptions might remain after teaching. Even when the teaching builds on prior knowledge of misconceptions, some misconceptions are very resistant to change. Furthermore, it is possible to learn some basic probability calculations in a short time, however, this does not mean the students really believe or have confidence in the results of their calculations and abandon their intuitions when they are contrary to the theory. They need experience in data collecting, data analysing, testing their judgement and adjusting their judgement and so on. The teaching of probability integrated with data analysis needs to be spread over time.

Third, at least for the short-term, the limitations on technology in many schools should not be considered as an impediment to introducing probability. The result of the teaching intervention in this study indicated that the computer demonstrations did not appear to play an important role in affecting students' change. To pool students' data or give ready data generated by computer are also effective approaches. This is likely to be particularly important in a situation where computers or software may not be readily available.

For Teachers

A particular problem with probabilistic reasoning is that it sometimes appears to be in conflict with causal, logical and deterministic thinking (Borovcnik and Peard, 1996). It is impossible to “prove” a theoretical probability by a trial or even a few trials. This causes difficulties for teaching and learning probability. Based on the findings of this study, the following four implications are suggested to teachers.

In order to save teaching time and in order to focus on strategies for solving examination problems, high school teachers in China are usually reluctant to organise classroom activities. However, students’ intuitions cannot be modified by verbal explanations alone (Fischbein & Gazit, 1984). The results of this study also show that for grade 12 students, although they had been exposed to a theoretical approach to probability, all the groups of conceptions were also observed. Consequently, the first suggestion is that teachers should create situations to encourage students to examine, modify, or correct their own beliefs about probability by the use of real data, activities and visual simulations (Hawkins & Kapadia, 1984, Garfield & Ahlgren, 1988, Konold, 1991, Shaughnessy, 1993).

Second, misconceptions exist in students’ minds and different students may have different misconceptions. The possible strategies used by school children and how they change over time as students develop is meaningful information that can be used in the teaching process. If teachers are aware of students’ misconceptions and take this into account in their teaching, the teaching should be more effective and efficient. They need to collect ongoing information on students' understanding of probability and adapt the activities to overcome these misconceptions. It is particularly important that teachers ask students to explain their reasons, both in writing and orally, since the same answer (whether correct or incorrect) could be the result of different reasoning (see Chapter 5).

Third, teachers need to help students to develop their understanding and lead them to give more sophisticated reasons. Even with students who gave correct answers and reasons, when analysed the reasons they gave were often not very sophisticated. Asking students to give reasons for their answers provides opportunities for students and teachers to discuss the reasons and improve students' reasoning: an important goal for teaching. The teaching intervention shows that with an appropriate approach it is possible to both overcome some misconceptions and increase the developmental level of students.

Fourth, students should have experience with multiple generators to model a probabilistic problem. It was found that the use of some misconceptions depended on data or context. Different generators such as dice, coins, spinners and playing cards were used in the teaching experiment. The result of the post-test shows that some students could identify multiple generators to model a probabilistic problem, while others still retain misconceptions within specific contexts or with particular data. If teachers use a variety of generators it should assist students developing their understanding of probability.

Generally speaking, if teaching is related to students' realities, including appropriate teaching approaches to disrupt their misconceptions and to develop their understanding structure, it is more likely to be meaningful and successful. If not, the students' subjective judgements, personal experiences and beliefs will still exist separately from what they are taught at school.

Further Research Topics

There are two main limitations of the main study. First, the sample of this study is a convenience sample. As discussed in chapter 3 it was not felt that this causes a problem with the validity of the results, but it does have some implications for further research. Specifically, the research should be replicated with other grades and in other areas of China.

Second, more than eighty items were used in this study. It enabled the researcher to obtain extensive information on the role that context and data played in eliciting students' misconceptions, but at the same time, for each item, only about 20 students in each grade wrote the item. This meant that the number of respondents for each item was limited. This limitation, together with some of the specific results from the questionnaires leads to a set of suggestions for further research.

Some particular misconceptions, such as, the compound approach, the different versions of equiprobability, and own calculation methods in chance comparison, are rarely studied systematically in the literature. In this study, they were reported and explained but further research on their justifications and effective teaching intervention strategies is needed.

Fourteen groups of misconceptions of probability were described in this study. In order to overcome them successfully, we need to know whether there are any relationships among the different groups of misconceptions or whether they could be organised by different schemata. Being aware of these relationships should help in finding effective teaching strategies.

It was found in this study that data and context played a role in some misconceptions. However, since for each item only about 60 students (across three grades and two streams) answered the item, the conclusions, such as, equiprobability are both data and context dependent, compound approach is data dependent, outcome approach is context dependent and chance cannot be measured mathematically is neither data nor context dependent, need to be examined in further research.

There were two major limitations associated with the teaching intervention. The first was the very short period, only 6 sessions, for the teaching intervention. Although the results indicated that even within this short period improvement could be made, a short intervention would seem to have limited its effectiveness. Can any gains in understanding be retained over time? How about a teaching intervention that focuses on the other (classical or subjective) definitions of probability? What are the effective teaching strategies for overcoming equiprobability and outcome approach? These are just some of the many questions that could be suggested for further research.

The second limitation concerns how the computer was used. Although the computer was used to demonstrate simulations in one of the classes during the teaching intervention, students did not have an opportunity to use the computer themselves. Instead, they watched the simulation process on the teacher's computer. This limited the freedom for student investigation and might also have limited any additional effects the computer might have had on developing students' understanding and helping them overcome misconceptions. The results of this study show that misconceptions can be overcome and students can increase their level of understanding without the computer. In fact, the results show that there was no difference whether or not the computer was used. However, further research on the role of the computer is needed. For example, would allowing the students to actually do the simulations be more effective than watching a teacher demonstration? Is the impact of the computer significant in a long-term teaching intervention? Which misconceptions might be most effectively overcome with the appropriate use of the computer?

As mentioned in chapter 1, students' understanding of probability is a potentially productive area of research. It is hoped this study becomes a reference for current curriculum reform in China, provides data on Chinese students' understanding of probability for other researchers, and finally, provides researchers with some interesting questions for future research.

Bibliography

Acredolo, C., O'Connor, J., Banks, L., & Horobin, K. (1989). Children's ability to make probability estimates: skills revealed through application of Anderson's functional measurement methodology. *Child Development*, 60, 933-945.

Andrich, D., Hess, R., & Ryan, J. M. (1998). *Applying measurement and assessment models for dichotomous and graded response formats using interactive software*. NCME Workshop, San Diego.

Australian Education Council. (1991). *A national statement on mathematics for Australian schools.* Carlton, VIC: Author.

Batanero, C., Serrano, L. & Garfield J. B. (1996). Heuristics and biases in secondary school students' reasoning about probability. In L. Puig, & Gutierrez (Eds.*), Proceedings of the 20th conferenceon the International Group for the Psychology of Mathematics Education, Volume2*, (pp. 51-59). University of Valencia.

Bell, A. W., Costello, J., & Küchemann, D. E. (Eds), (1983). *A Review of Research in Mathematics Education*. NFER-NELSON Publishing Company Ltd.

Benson, C. T. & Jones, G. A. (1999). Assessing students' thinking in modeling probability contexts. *The Mathematics Educator*, 4(2), 1-21.

Bentz, H. & Borovcnik, M. (1985). Probability and statistics. In Bell, A., Low, B. and Kilpatrick, J. (Eds), *Theory, Research & Practice in Mathematical Education*. Sell Centre for Mathematical Education, University of Nottingham (pp. 261-337).

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic Press.

Biggs, J. B., & Collis, K. F. (1991). Multimodal learning and the quality of intelligent behavior. In H. A. H. Rowe. (Ed.), *Intelligence: Reconceptualisation and measurement* (pp. 57-76). Hillsdale, N.J.: Lawrence Erlbaum.

Borovcnik, M., & Bentz, H. J. (1991). Empirical research in understanding probability. In R. Kapadia & M. Borovcnik (Eds.), *Chance encounters: Probability in education* (pp. 73-105). Dordrecht, The Netherlands: Kluwer.

Borovcnik, M., & Peard, R. (1996). Probability. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick & C. Laborde (Eds.), *International handbook of mathematics education* (pp. 239-287). Dordrecht, The Netherlands: Kluwer.

Brown, C. A., & Silver, E. A. (1989). Discrete mathematics. In M. M. Lindquist, (Ed.), *Results from the fourth mathematics assessment of the national assessment of educational progress* (pp. 19-27). Reston, VA: The Council.

Cañizares, M.J., & Batanero, C. (1998). A study on the stability of equiprobability bias in 10-14 year-old children. In L. Pereira-Mendoza (Eds.), *Statistical education – Expanding the network. Proceedings of the Fifth International Conference On Teaching Statistics, Volume 2*. (pp. 1447-1448). Voorburg, The Netherlands: International Statistical Institute.

Carpenter, T. P., Corbitt, M. K., Kepner, JR. H. S., Lindquist, M. M., and Reys, R. E. (1981). What are the chances of your students knowing probability? *Mathematics Teacher*, 74, 342-345.

Chan, W. S. (1997). *16 to 18 year old students' errors and misconceptions in learning probability.* Unpublished thesis (M. Ed.). Nanyang Technological University, Singapore.

Chick, H. L., Watson, J. M. & Collis, K. F. (1988). Using the SOLO taxonomy for error analysis in mathematics. *Research in Mathematics Education in Australia*, 1988(6), 34-46.

Collis, K. F., Romberg, T. A., & Jurdak, M. E. (1986). A technique for assessing mathematical problem-solving ability. *Journal for Research in Mathematics Education*, 17, 206-221.

Collis, K. F., & Biggs, J. B. (1991). Developmental determinants of qualitative aspects of school learning. In G. Evans (Ed.), *Learning and teaching cognitive skills.* (pp. 185-207). Melbourne: Australian Council for Educational Research.

Collis, K. F. (1992). Curriculum and assessment: a basic cognitive model. In G. Leder (Ed.), *Assessment and learning of mathematics* (pp. 24-45). Hawthorn, Vic.: Australian Council for Educational Research.

Cox, C., & Mouw, J. T. (1992). Disruption of the representativeness heuristic: Can we be perturbed into using correct probabilistic reasoning? *Educational Studies in Mathematics*, 23, 163-178.

DeGroot, M. H. (1986). *Probability and statistics*. Reading, Mass.: Addison-Wesley Publishing Company.

English, L. D. (1993). Children's strategies for solving two- and three-dimensional combinatorial problems. *Journal for Research in Mathematics Education*, 24(3), 255-273.

Ernest, P. (1984). Introducing the concept of probability. *Mathematics Teacher*, 77, 524-525.

Falk, R., Falk, R., & Levin, I. (1980). A potential for learning probability in young children. *Educational Studies in Mathematics*, 11, 181-204.

Falk, R. (1981). The perception of randomness. In C. Comiti (Ed.) *Proceedings of the Fifth International Conference for the Psychology of Mathematics Education*. (pp. 222-229). Grenoble, France.

Falk, R. (1983). Children's choice behaviour in probabilistic situations. In D. R. Grey, P. Holmes, V. Barnett, & G. M. Constable (Eds.), *Proceedings of the First International Conference On Teaching Statistics, Volume* 2. (pp. 714-726). Sheffield, U.K.: Teaching Statistics Trust.

Fischbein, E. (1975). *The intuitive sources of probabilistic thinking in children.* Dordrecht: Reidel.

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

Fischbein, E., Nello, M. S., & Marino, M. S. (1991). Factors affecting probabilistic judgments in children and adolescents. *Educational Studies in Mathematics*, 22, 523-549.

Fischbein, E., & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research in Mathematics Education*, 28(1), 96-105.

Freudenthhal, H. (1973). *Mathematics as an educational task.* Dordrecht: Reidel.

Garfield, J. B., & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: Implications for research. *Journal for Research in Mathematics Education,* 19, 44-63.

Garfield, J. B. (1995). Reflections on the past 15 years. *Teaching Statistics,* 17(2), 77-78.

Garfield, J. B. (1998). The statistical reasoning assessment: Development and validation of a research tool. In L. Pereira-Mendoza (Eds.), *Statistical education – Expanding the network. Proceedings of the Fifth International Conference On Teaching Statistics, Volume 2*. (pp. 781-786). Voorburg, The Netherlands: International Statistical Institute.

Green, D. R. (1982). *Probability concepts in 11-16 year old pupils.* 2nd edition. Centre for Advancement of Mathematical Education in Technology, University of Technology, Loughborough.

Green, D. R. (1983a). School pupils' probability concepts. *Teaching Statistics*, 5(2), 34-42.

Green, D. R. (1983b). A Survey of probability concepts in 3000 pupils aged 11-16 years. In D. R. Grey, P. Holmes, V. Barnett & G. M. Constable (Eds.), *Proceedings of the First International Conference on Teaching Statistics, Volume 2* (pp. 766-783). Sheffield, U.K.: Teaching Statistics Trust.

Green, D. R. (1983c). Shaking a six. *Mathematics in Schools*, 12(5), 29-32.

Hawkins, A., & Kapadia, R. (1984). Children's conceptions of probability --- A psychological and pedagogical review. *Educational Studies in Mathematics*, 15, 349-377.

Hawkins, A., Jollifee, F., & Glickman, L. (1992). *Teaching statistical concepts*. London: Longman.

Hoemann, H. W., & Ross, B. M. (1982). Children's concepts of chance and probability. In C. J. Brainerd (Ed.), *Children's logical and mathematics cognition*. (pp. 93-121). New York: Springer-Verlag.

Huff, D. (1959). *How to take a chance?* New York: W. W. Norton.

Interactive Mathematics Program. (1998). *Interactive Mathematics Program. (Integrated High School Mathematics).* Berkeley, CA: Key Curriculum Press.

Johson, R. (1992). *Elementary Statistics*. Boston, Massachusetts: PWS-KENT Publishing Company.

Jones, G. A., Langrall, C. W. Thornton, C. A., & Mogill, A. T. (1997). A framework for assessing and nurturing young children's thinking in probability. *Educational Studies in Mathematics*, 32, 101-125.

Jones, G. A., Langrall, C. W., Thornton, C. A., & Mogill, A. T. (1999). Students' probabilistic thinking in instruction. *Journal for Research in Mathematics Education*, 30, 487-519.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. In D. Kahneman (Eds.), *Judgment under uncertainty: Heuristics and biases*. (pp.32-47). Cambridge: Cambridge University Press, Cambridge, U.K.

Kapadia, R. (1986). Didactical phenomenology of probability. In R. Davidson & J. Swift (Eds.), *The Proceedings of the Second International Conference on Teaching Statistics*. (pp. 260-264). Victoria, B. C.: University of Victoria.

Kapadia, R., & Borovcnik, M. (1991). *Chance encounters: Probability in education.* Dordrecht: Kluwer.

Karplus, R., Pulos, S., & Stage, E. K. (1983). Proportional reasoning of early adolescents. In R. Lesh & M. Landau (Eds.), *Acquisition of mathematics concepts and processes.* (pp. 45-90). London: Academic Press.

Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction*, 6, 59-98.

Konold, C. (1991). Understanding students' beliefs about probability. In E. von Glasersfeld (Ed.), *Radical constructivism in mathematics education*. (pp. 139-156). Dordrecht: Kluwer.

Konold, C. (1995). Issues in assessing conceptual understanding in probability and statistics. *Journal of Statistics Education (online)* 3(1).

Konold, C., Pollatsek, A., Well, A.D., Lohmeier, J., & Lipson, A. (1993). Inconsistencies in students' reasoning about probability. *Journal for Research in Mathematics Education*, 24, 392-414.

Lam, P., & Foong, Y. Y. (1996). *Rasch analysis of math SOLO taxonomy levels using hierarchical items in testlets*. ERIC ED398271.

Lecoutre, M. P., Durand, J., & Cordier, J. (1990). A study of two biases in probabilistic judgments: representativeness and equiprobability. In J.-P. Caverni, J.-M. Fabre & M. Gonzalès (Eds.), *Cognitive Biases*. (pp. 563-575). North Holland: Elsevier Science Publishers.

Lecoutre, M. P. (1992). Cognitive models and problem spaces in "purely random" situations. *Educational Studies in Mathematics*, 23, 557-568.

Lidster, S. T., Pereira-Mendoza, L., Watson, J. M., & Collis, K. F. (1995). What's fair for Grade 6? Presented at the annual conference of the Australian Association for Research in Education, Hobart, November, 1995.

Lovitt, C., & Lowe, I. (1993). *Chance and Data Investigations.* Volumes 1 & 2, Carlton, VIC: Curriculum Corporation.

Madsen, R. (1995). Secondary students' concepts of probability. *Teaching Statistics*, 17(3), 90-93.

Masters, G. N. (1988). Measurement models for ordered response categories. In R. Langeheine & J. Rost (Ed.). Latent trait and latent class models (pp.11-28). London: Plenum Press.

Metz, K. E. (1997). Dimensions in the assessment of students' understanding and application of chance. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 223-238). Amsterdam: IOS Press and The International Statistical Institute.

Ministry of Education. (1992). *Mathematics in the New Zealand Curriculum.* Wellington, New Zealand: Author.

Monks, A. R. (1985). Equally likely. *Teaching Statistics*, 7(3), 66-69.

Moritz, J. B., Watson, J. M., & Pereira-Mendoza, L. (1996). The language of statistical understanding: an investigation in two countries. Paper presented at the annual conference of the Australian Association for Research in Education, Singapore, November, 1996.

Moritz, J. B., Watson, J. M., & Collis, K. F. (1996). Odds: Chance measurement in three contexts. In P. C. Clarkson (Ed.), *Technology in Mathematics Education* (pp. 390-397). Melbourne: Mathematics Education Research Group of Australia.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: The Council.

National Council of Teachers of Mathematics. (1991). *Dealing with data and chance.* Reston, VA: The Council.

National Council of Teachers of Mathematics. (1992). *Data analysis and statistics across the curriculum.* Reston, VA: The Council.

National Research Council. (1989). *Everybody counts: A report to the nation on the future of mathematics education.* Washington, DC: National Academy Press.

Nemetz, T. (1997). State of the art of teaching probability at the secondary level. In B. Phillips (Ed.), Papers in Statistical Education presented at ICME-8 (pp. 75-86). Swinburne (Australia): Swinburne, University of Technology and IASE.

Newman, C. M., Obremski, T. E., & Scheaffer, R. L. (1987). *Exploring probability.* Palo Alto: Dale Seymour.

Noelting, G. (1980a). The development of proportional reasoning and the ratio concept. Part I --- Differentiation of stages. *Educational Studies in Mathematics*, 11, 217-253.

Noelting, G. (1980b). The development of proportional reasoning and the ratio concept. Part II --- Problem-structure at successive stages: Problem-solving strategies and the mechanism of adaptive restructuring. *Educational Studies in Mathematics*, 11, 331-363.

Orton, A. (1994). Learning mathematics: Implications for teaching. In A. Orton & G. Wain (Eds.), *Issues in teaching mathematics*. London: Cassell.

Piaget, J., & Inhelder, B. (1975). *The origin of the idea of chance in children*. (L. Leake, Jr., P. Burrell & H. D. Fishbein, Trans.). New York: Norton. (Original work published 1951).

Rogers, L. (1992). Early years mathematics: children, teachers and assessment. In Gilah Leder (Ed.), *Assessment and learning of mathematics* (pp. 223-248). Hawthorn, Vic.: Australian Council for Educational Research.

Rowntree, D. (1984). *Probability without tears.* New York: Scribner's.

Rasch Unidimensional Measurement Models. (2000) *v3.0*. [Computer programme]. Rumm Laboratory, Australia.

Study group of school mathematics for the twenty-first century in China, (1995). *School mathematics for the twenty-first century in China.* Beijing: Beijing Normal University.

Shafer, G. (1992). What is probability? In D. C. Hoaglin & D. S. Moore (Eds.), *Perspectives on contemporary statistics.* (pp. 93-105). Washington, D. C.: Mathematical Association of America.

Shaughnessy, J. M. (1977). Misconceptions of probability: an experiment with a small-group, activity-based, model building approach to introductory probability at the college level. *Educational Studies in Mathematics*, 8, 285-316.

Shaughnessy, J. M. (1983). Misconceptions of probability: systematic and otherwise: Teaching probability and statistics so as to overcome some misconceptions. In D. R. Grey, P. Holmes, V. Barnett & G. M. Constable (Eds.), *Proceedings of the First International Conference on Teaching Statistics, Volume 2*. (pp. 784-801). Sheffield, U.K.: Teaching Statistics Trust.

Shaughnessy, J. M. (1992). Research in probability and statistics: reflections and directions. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning*. (pp. 465-494). New York: Macmillan.

Shaughnessy, J. M. (1993). Probability and statistics. *The Mathematics Teacher*, 86, 244-248.

Silver, E. A. (1987). Foundations of cognitive theory and research for mathematics problem-solving. In A. H. Shoenfeld (Ed.), *Cognitive science and mathematics education*. (pp. 33-60). New Jersey: Lawrence Erlbaum Associates.

Simon, J. L. (1992). *Resampling: The new statistics*. Belmont, CA: Duxbury Press.

Singer, J., & Resnick, L. (1992). Representations of proportional relationships: Are children part-part or part-whole reasoners? *Educational Studies in Mathematics*, 23, 231-246.

Smith. P. K., & Cowie. H. (1988). *Understanding children's development*. Basil Blackwell Ltd, UK.

Strait, P. T. (1983). *A first course in probability and statistics with applications*. New York: Harcourt Brace Jovanovich.

Strauss, S., & Bichler, E. (1988). The development of children's concepts of the arithmetic average. *Journal for Research in Mathematics Education*, 19(1) 64-80.

Tangible MATH 3.04 -- Probability Constructor (1998). [Computer programme]. LOGAL software Inc. U.S.

Truran, J. (1985). Children's understanding of symmetry. *Teaching Statistics*, 7(3), 69-74.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 3-20). Cambridge: Cambridge University Press.

Watson, J. M., Campbell, K. J., & Collis, K. F. (1993). Multimodal functioning in understanding fractions. *Journal of Mathematical Behaviour,* 12, 45-62.

Watson, J. M. (1995). Probability and statistics: an overview. In L. Grimison & J. Pegg (Eds.), *Teaching secondary school mathematics: theory in practice*. (pp.120-139). Sydney: Harcourt Brace.

Watson, J. M., Collis, K. F., & Moritz, J. B. (1997). The development of chance measurement. *Mathematics Education Research Journal*, 9, 60-82.

Watson, J. M., & Moritz, J. B. (1998). Longitudinal development of chance measurement. Manuscript.

Watson, J. M., & Moritz, J. B. (1999). Developing concepts of sampling. *Journal for Research in Mathematics Education*, 31(1), 44-70.

Way, J. (1996). A study of children's probability judgments. In P. C. Clarkson (Ed.), *Technology in mathematics education, Proceedings of the 19th annual conference of the Mathematics Education Research Group of Australasia* (pp. 596-603). Melbourne: Mathematics Education Research Group of Australasia.

Way, J. (1998). Young children's probabilistic thinking. In L. Pereira-Mendoza (Eds.), *Statistical education – Expanding the network. Proceedings of the Fifth International Conference on Teaching Statistics, Volume 2*. (pp. 765-771). Voorburg, The Netherlands: International Statistical Institute.

Williams, J. S., & Amir, G. S. (1995). 11-12 year old children's informal knowledge and its influence on their formal probabilistic reasoning. ERIC ED 387256.

Zhang, F. S. Zhou, Y. X., & Zhao, X. P. (1998). Close to real life, appreciate application and emphasise involvement. *Journal of Mathematics Teaching and Learning (Chinese)* 159, 1-5.

Appendix A

Items for the Main Study
(English and Chinese Versions)

The following are some general notes on the items:

1. This is a summary of all the items used in the main study. As explained in the thesis different combinations were used in different questionnaires.

2. For each item students were asked to indicate their confidence. This was presented at the end of each item in the following format.

no confidence	little confidence	a lot of confidence	full confidence
☐	☐	☐	☐

3. For each item students were asked to explain their answer.

4. For category I items I1 - I4 each part was presented separately. At the end of each part students were asked to put a tick in the appropriate box.

impossible	possible	certain
☐	☐	☐

5. For category III items III1 - III4, parts (1) and (2) were always used together.

6. Since the majority of the students had not been taught probability, in the items the word outcome rather than event was used, even in situations where event is the usual term.

7. The questionnaires were administered in Chinese so the items were written to be as clear as possible in Chinese. In translating them back into English the final English version may appear slightly awkward due to differences between the languages, both in wording and structure. The translation was done in this manner to ensure that the meaning was not changed due to the translation.

Category I items:

I1 A six-sided normal die is rolled once. Please indicate whether the following outcome is impossible, possible or certain to happen. Tick where appropriate:
(1) the number rolled is an even number.
(2) the number rolled is smaller than 7.
(3) the number rolled is bigger than 6.
(4) the number rolled is 2.
(5) the number rolled is 6.
(6) the number rolled is not 6.

I2 There is a pack of cards without picture cards. One card is drawn from the pack of shuffled cards. The Ace is worth 1. Please indicate whether the following outcome is impossible, possible or certain to happen. Tick where appropriate:
(1) the card drawn is heart or diamond.
(2) the card drawn is smaller than a Jack.
(3) the card drawn is bigger than a 10.
(4) the card drawn is the 2 of diamonds.
(5) the card drawn is a 10.
(6) the card drawn is not a 10.

I3 Three six-sided normal dice are rolled once. Please indicate whether the following outcome is impossible, possible or certain to happen. Tick where appropriate:
(1) all three numbers rolled are even numbers.
(2) all three numbers rolled are smaller than 7.
(3) all three numbers rolled are bigger than 6.
(4) all three numbers rolled are 2.
(5) all three numbers rolled are 6.
(6) None of the three numbers rolled is 6.

I4 There are three packs of cards without picture cards. One card is drawn from each pack of shuffled cards. The Ace is worth 1. Please indicate whether the following outcome is impossible, possible or certain to happen. Tick where appropriate:
(1) all three cards drawn are hearts or diamonds.
(2) all three cards drawn are smaller than a Jack.
(3) all three cards drawn are bigger than a 10.
(4) the three cards drawn are the 2 of diamonds, the 5 of diamonds and the 8 of spades.
(5) all three cards drawn are 2 of diamonds.
(6) all three cards drawn are different in both number and suit.

Category II items:

II1(1) Before a match, the coach predicted that "based on the information I have our chance of winning the game is 80%." Which of the following has the closest meaning to the "chance of winning the game is 80%"?

a) his team will certainly win this match
b) his team will certainly lose this match
c) suppose that the match could be repeated 10 times, his team wins around 8 out of the 10 matches
d) suppose that the match could be repeated 10 times, his team wins exactly 8 out of the 10 matches

II1(2) Before a match, the coach predicted that "based on the information I have our chance of winning the game is 50%." Which of the following has the closest meaning to the "chance of winning the game is 50%"?

a) his team may win or may lose. He doesn't really know the result of this match
b) suppose that the match could be repeated 10 times, his team wins around 5 out of the 10 matches
c) suppose that the match could be repeated 100 times, his team wins exactly 50 out of the 100 matches
d) suppose that the match could be repeated 100 times, his team wins around 50 out of the 100 matches

II1(3) Before a match, the coach predicted that "based on the information I have, our chance of winning the game is 30%." Which of the following has the closest meaning to the "chance of winning the game is 30%"?

a) his team will certainly win this match
b) his team will certainly lose this match
c) suppose that the match could be repeated 10 times, his team wins around 3 out of the 10 matches
d) suppose that the match could be repeated 10 times, his team wins exactly 3 out of the 10 matches

II1(4) Before a match, the coach predicted that "based on the information I have, our chance of winning the game is 80%." The game is played and the team loses. Do you think the coach's prediction is accurate or not?

II1(5) Before a match, the coach predicted that "based on the information I have, our chance of winning the game is 50%." The game is played and the team loses. Do you think the coach's prediction is accurate or not?

II1(6) Before a match, the coach predicted that "based on the information I have, our chance of winning the game is 30%." The game is played and the team loses. Do you think the coach's prediction is accurate or not?

II1(7) Before a match, the coach predicted that "based on the information I have, our chance of winning the game is 80%." Here are five situations. Comparatively speaking, in which situation can the coach's prediction be considered very accurate?

a) his team really wins this match
b) his team really loses this match
c) suppose that the match could be repeated 10 times, his team wins all the 10 matches
d) suppose that the match could be repeated 10 times, his team wins 9 out of the 10 matches
e) suppose that the match could be repeated 10 times, his team wins 8 out of the 10 matches

II1(8) Before a match, the coach predicted that "based on the information I have, our chance of winning the game is 50%." Here are five situations. Comparatively speaking, in which situation can the coach's prediction be considered very accurate?

a) his team really wins this match
b) his team really loses this match
c) suppose that the match could be repeated 10 times, his team wins 5 out of the 10 matches
d) suppose that the match could be repeated 10 times, his team wins 3 out of the 10 matches
e) suppose that the match could be repeated 100 times, his team wins 50 out of the 100 matches

II1(9) Before a match, the coach predicted that "based on the information I have, our chance of winning the game is 30%." Here are five situations. Comparatively speaking, in which situation can the coach's prediction be considered very accurate?

a) his team really wins this match
b) his team really loses this match
c) suppose that the match could be repeated 10 times, his team loses all the 10 matches
d) suppose that the match could be repeated 10 times, his team wins 1 out of the 10 matches
e) suppose that the match could be repeated 10 times, his team wins 3 out of the 10 matches

II2(1) A mathematician filled a bag with black and white marbles. He didn't really know how many black marbles and white marbles were in the bag. After mixing them thoroughly, he took a look and predicted that "if I pull out a marble from the bag without looking, the chance that it will happen to be white is 80%." Which of the following has the closest meaning to "the chance that it will happen to be white is 80%"?

a) the marble pulled out will certainly be white
b) the marble pulled out will certainly be black
c) suppose that the game is repeated 10 times, white marbles are pulled out around 8 times
d) suppose that the game is repeated 10 times, white marbles are pulled out exactly 8 times

II2(2) A mathematician filled a bag with black and white marbles. He didn't really know how many black marbles and white marbles were in the bag. After mixing them thoroughly, he took a look and predicted that "if I pull out a marble from the bag without looking, the chance that it will happen to be white is 50%." Which of the following has the closest meaning to "the chance that it will happen to be white is 50%"?

a) the marble pulled out may be white or may be black. The mathematician doesn't really know what the result will be
b) suppose that the game is repeated 10 times, white marbles are pulled out around 5 times
c) suppose that the game is repeated 100 times, white marbles are pulled out exactly 50 times
d) suppose that the game is repeated 100 times, white marbles are pulled out around 50 times

II2(3) A mathematician filled a bag with black and white marbles. He didn't really know how many black marbles and white marbles were in the bag. After mixing them thoroughly, he took a look and predicted that "if I pull out a marble from the bag without looking, the chance that it will happen to be white is 30%." Which of the following has the closest meaning to "the chance that it will happen to be white is 30%"?

a) the marble pulled out will certainly be white
b) the marble pulled out will certainly be black
c) suppose that the game is repeated 10 times, white marbles are pulled out around 3 times
d) suppose that the game is repeated 10 times, white marbles are pulled out exactly 3 times

II2(4) A mathematician filled a bag with black and white marbles. He didn't really know how many black marbles and white marbles were in the bag. After mixing them thoroughly, he took a look and predicted that "if I pull out a marble from the bag without looking, the chance that it will happen to be white is 80%." He pulls out a marble. The marble is white. Do you think the mathematician's prediction is accurate or not?

II2(5) A mathematician filled a bag with black and white marbles. He didn't really know how many black marbles and white marbles were in the bag. After mixing them thoroughly, he took a look and predicted that "if I pull out a marble from the bag without looking, the chance that it will happen to be white is 50%." He pulls out a marble. The marble is white. Do you think the mathematician's prediction is accurate or not?

II2(6) A mathematician filled a bag with black and white marbles. He didn't really know how many black marbles and white marbles were in the bag. After mixing them thoroughly, he took a look and predicted that "if I pull out a marble from the bag without looking, the chance that it will happen to be white is 30%." He pulls out a marble. The marble is white. Do you think the mathematician's prediction is accurate or not?

II2(7) A mathematician filled a bag with black and white marbles. He didn't really know how many black marbles and white marbles were in the bag. After mixing them thoroughly, he took a look and predicted that "if I pull out a marble from the bag without looking, the chance that it will happen to be white is 80%." Here are five situations. Comparatively speaking, in which situation can the mathematician's prediction be considered very accurate?

a) pulling out a marble and it happens to be a white marble
b) pulling out a marble and it happens to be a black marble
c) suppose that the game is repeated 10 times, white marbles are pulled out 10 times
d) suppose that the game is repeated 10 times, white marbles are pulled out 9 times
e) suppose that the game is repeated 10 times, white marbles are pulled out 8 times

II2(8) A mathematician filled a bag with black and white marbles. He didn't really know how many black marbles and white marbles were in the bag. After mixing them thoroughly, he took a look and predicted that "if I pull out a marble from the bag without looking, the chance that it will happen to be white is 50%." Here are five situations. Comparatively speaking, in which situation can the mathematician's prediction be considered very accurate?

a) pulling out a marble and it happens to be a white marble
b) pulling out a marble and it happens to be a black marble
c) suppose that the game is repeated 10 times, white marbles are pulled out 5 times
d) suppose that the game is repeated 10 times, white marbles are pulled out 3 times
e) suppose that the game is repeated 100 times, white marbles are pulled out 50 times

II2(9) A mathematician filled a bag with black and white marbles. He didn't really know how many black marbles and white marbles were in the bag. After mixing them thoroughly, he took a look and predicted that "if I pull out a marble from the bag without looking, the chance that it will happen to be white is 30%." Here are five situations. Comparatively speaking, in which situation can the mathematician's prediction be considered very accurate?

a) pulling out a marble and it happens to be a white marble
b) pulling out a marble and it happens to be a black marble
c) suppose that the game is repeated 10 times, black marbles are pulled out 10 times
d) suppose that the game is repeated 10 times, white marbles are pulled out 1 time
e) suppose that the game is repeated 10 times, white marbles are pulled out 3 times

II3(1) A weather forecaster said that "tomorrow there is an 80% chance that it will rain." Which of the following has the closest meaning to "tomorrow there is an 80% chance that it will rain"?

a) it will certainly rain tomorrow
b) it will certainly not rain tomorrow
c) suppose that there were 10 days in a year the forecaster said that "tomorrow there is an 80% chance that it will rain", and on around 8 out of the10 days it rains the next day
d) suppose that there were 10 days in a year the forecaster said that "tomorrow there is an 80% chance that it will rain", and on exactly 8 out of the10 days it rains the next day

II3(2) A weather forecaster said that "tomorrow there is a 50% chance that it will rain." Which of the following has the closest meaning to "tomorrow there is a 50% chance that it will rain"?

a) it may or may not rain tomorrow. The forecaster doesn't really know what the result will be
b) suppose that there were 10 days in a year the forecaster said that "tomorrow there is a 50% chance that it will rain", and on around 5 out of the 10 days it rains the next day
c) suppose that there were 100 days in a year the forecaster said that "tomorrow there is a 50% chance that it will rain", and on exactly 50 out of the 100 days it rains the next day
d) suppose that there were 100 days in a year the forecaster said that "tomorrow there is a 50% chance that it will rain", and on around 50 out of the 100 days it rains the next day

II3(3) A weather forecaster said that "tomorrow there is a 30% chance that it will rain." Which of the following has the closest meaning to “tomorrow there is a 30% chance that it will rain”?

a) it will certainly rain tomorrow
b) it will certainly not rain tomorrow.
c) suppose that there were 10 days in a year the forecaster said that “tomorrow there is a 30% chance that it will rain”, and on around 3 out of the10 days it rains the next day
d) suppose that there were 10 days in a year the forecaster said that “tomorrow there is a 30% chance that it will rain”, and on exactly 3 out of the 10 days it rains the next day

II3(4) A weather forecaster said that "tomorrow there is an 80% chance that it will rain." The next day it doesn’t rain. Do you think the forecaster’s prediction is accurate or not?

II3(5) A weather forecaster said that “tomorrow there is a 50% chance that it will rain.” The next day it doesn’t rain. Do you think the forecaster’s prediction is accurate or not?

II3(6) A weather forecaster said that “tomorrow there is a 30% chance that it will rain.” The next day it doesn’t rain. Do you think the forecaster’s prediction is accurate or not?

II3(7) A weather forecaster said that "tomorrow there is an 80% chance that it will rain." Here are five situations. Comparatively speaking, in which situation can the forecaster’s prediction be considered very accurate?

a) it really rains the next day
b) it really doesn’t rain the next day
c) suppose that there were 10 days in a year the forecaster said that “tomorrow there is an 80% chance that it will rain”, and on all the 10 days it rains the next day
d) suppose that there were 10 days in a year the forecaster said that “tomorrow there is an 80% chance that it will rain”, and on 9 out of the 10 days it rains the next day
e) suppose that there were 10 days in a year the forecaster said that “tomorrow there is an 80% chance that it will rain”, and on 8 out of the 10 days it rains the next day

II3(8) A weather forecaster said that "tomorrow there is a 50% chance that it will rain." Here are five situations. Comparatively speaking, in which situation can the forecaster's prediction be considered very accurate?

a) it really rains the next day
b) it really doesn't rain the next day
c) suppose that there were 10 days in a year the forecaster said that "tomorrow there is a 50% chance that it will rain", and on 5 out of the 10 days it rains the next day
d) suppose that there were 10 days in a year the forecaster said that "tomorrow there is a 50% chance that it will rain", and on 3 out of the 10 days it rains the next day
e) suppose that there were 100 days in a year the forecaster said that "tomorrow there is a 50% chance that it will rain", and on 50 out of the 100 days it rains the next day

II3(9) A weather forecaster said that "tomorrow there is a 30% chance that it will rain." Here are five situations. Comparatively speaking, in which situation below the forecaster's prediction can be considered very accurate?

a) it really rains the next day
b) it really doesn't rain the next day
c) suppose that there were 10 days in a year the forecaster said that "tomorrow there is a 30% chance that it will rain", and on all the 10 days it does not rain the next day
d) suppose that there were 10 days in a year the forecaster said that "tomorrow there is a 30% chance that it will rain", and on 1 out of the 10 days it rains the next day
e) suppose that there were 10 days in a year the weather forecaster said that "tomorrow there is a 30% chance that it will rain", and on 3 out of the 10 days it rains the next day

Category III items:

III1 A class has 20 girls and 22 boys in it. Each pupil's name is written on a piece of paper and all the names are put into a box and mixed thoroughly.

(1) The teacher picks 1 name out of the box casually without looking. Which statement below is correct?

a) it is more likely to pick out a boy's than a girl's name
b) it is less likely to pick out a boy's than a girl's name
c) it is equally likely to pick out a boy's as a girl's name
d) it is impossible to compare the likelihood of the two outcomes

(2) Return the name into the box and mix thoroughly again. Now, the teacher picks 6 names out of the box casually without looking. The 6 names are 2 girls and 4 boys. He placed the 6 slips on the table, closes his eyes and does the 7^{th} draw from among the rest of the slips in the box. Which statement below is correct?
   a) it is more likely to pick out a boy's than a girl's name this time
   b) it is less likely to pick out a boy's than a girl's name this time
   c) it is equally likely to pick out a boy's as a girl's name this time
   d) it is impossible to compare the likelihood of the two outcomes

(3) The teacher picks 6 names out of the box casually without looking. The 6 names are 3 girls and 3 boys. He placed the 6 slips on the table, closes his eyes and does the 7^{th} draw from among the rest of the slips in the box. Which statement below is correct?
   a) it is more likely to pick out a boy's than a girl's name this time
   b) it is less likely to pick out a boy's than a girl's name this time
   c) it is equally likely to pick out a boy's as a girl's name this time
   d) it is impossible to compare the likelihood of the two outcomes

(4) Return all the names into the box and mix thoroughly again. Now, the teacher picks 6 names out of the box casually without looking. The 6 names are all boys. He placed the 6 slips on the table, closes his eyes and does the 7^{th} draw among the rest of the slips in the box. Which statement below is correct?
   a) it is more likely to pick out a boy's than a girl's name this time
   b) it is less likely to pick out a boy's than a girl's name this time
   c) it is equally likely to pick out a boy's as a girl's name this time
   d) it is impossible to compare the likelihood of the two outcomes

III2 A class has 5 girls and 27 boys in it. Each pupil's name is written on a piece of paper and all the names are put into a box and mixed thoroughly.

(1) The teacher picks 1 name out of the box casually without looking. Which statement below is correct?
   a) it is more likely to pick out a boy's than a girl's name
   b) it is less likely to pick out a boy's than a girl's name
   c) it is equally likely to pick out a boy's as a girl's name
   d) it is impossible to compare the likelihood of the two outcomes

(2) Return the name into the box and mix thoroughly again. Now, the teacher picks 6 names out of the box casually without looking. The 6 names are 2 girls and 4 boys. He placed the 6 slips on the table, closes his eyes and does the 7^{th} draw from among the rest of the slips in the box. Which statement below is correct?
   a) it is more likely to pick out a boy's than a girl's name this time
   b) it is less likely to pick out a boy's than a girl's name this time
   c) it is equally likely to pick out a boy's as a girl's name this time
   d) it is impossible to compare the likelihood of the two outcomes

(3) The teacher picks 6 names out of the box casually without looking. The 6 names are 3 girls and 3 boys. He placed the 6 slips on the table, closes his eyes and does the 7^{th} draw from among the rest of the slips in the box. Which statement below is correct?
   a) it is more likely to pick out a boy's than a girl's name this time
   b) it is less likely to pick out a boy's than a girl's name this time
   c) it is equally likely to pick out a boy's as a girl's name this time
   d) it is impossible to compare the likelihood of the two outcomes

(4) Return all the names into the box and mix thoroughly again. Now, the teacher picks 6 names out of the box casually without looking. The 6 names are all boys. He placed the 6 slips on the table, closes his eyes and does the 7th draw among the rest of the slips in the box. Which statement below is correct?

a) it is more likely to pick out a boy's than a girl's name this time
b) it is less likely to pick out a boy's than a girl's name this time
c) it is equally likely to pick out a boy's as a girl's name this time
d) it is impossible to compare the likelihood of the two outcomes

III3 A school has 400 girls and 440 boys in it. Each pupil's name is written on a piece of paper and all the names are put into a box and mixed thoroughly.

(1) The principal picks 1 name out of the box casually without looking. Which statement below is correct?

a) it is more likely to pick out a boy's than a girl's name
b) it is less likely to pick out a boy's than a girl's name
c) it is equally likely to pick out a boy's as a girl's name
d) it is impossible to compare the likelihood of the two outcomes

(2) Return the name into the box and mix thoroughly again. Now, the principal picks 70 names out of the box casually without looking. The 70 names are 15 girls and 55 boys. He placed the 70 slips on the table, closes his eyes and does the 71st draw from among the rest of the slips in the box. Which statement below is correct?

a) it is more likely to pick out a boy's than a girl's name this time
b) it is less likely to pick out a boy's than a girl's name this time
c) it is equally likely to pick out a boy's as a girl's name this time
d) it is impossible to compare the likelihood of the two outcomes

(3) The principal picks 70 names out of the box casually without looking. The 70 names are 35 girls and 35 boys. He placed the 70 slips on the table, closes his eyes and does the 71st draw from among the rest of the slips in the box. Which statement below is correct?

a) it is more likely to pick out a boy's than a girl's name this time

b) it is less likely to pick out a boy's than a girl's name this time

c) it is equally likely to pick out a boy's as a girl's name this time

d) it is impossible to compare the likelihood of the two outcomes

(4) Return all the names into the box and mix thoroughly again. Now, the principal picks 70 names out of the box casually without looking. The 70 names are all boys. He placed the 70 slips on the table, closes his eyes and does the 71st draw from among the rest of the slips in the box. Which statement below is correct?

a) it is more likely to pick out a boy's than a girl's name this time

b) it is less likely to pick out a boy's than a girl's name this time

c) it is equally likely to pick out a boy's as a girl's name this time

d) it is impossible to compare the likelihood of the two outcomes

III4 A school has 200 girls and 1000 boys in it. Each pupil's name is written on a piece of paper and all the names are put into a box and mixed thoroughly.

(1) The principal picks 1 name out of the box casually without looking. Which statement below is correct?

a) it is more likely to pick out a boy's than a girl's name

b) it is less likely to pick out a boy's than a girl's name

c) it is equally likely to pick out a boy's as a girl's name

d) it is impossible to compare the likelihood of the two outcomes

(2) Return the name into the box and mix thoroughly again. Now, the principal picks 70 names out of the box casually without looking. The 70 names are 23 girls and 47 boys. He placed the 70 slips on the table, closes his eyes and does the 71^{st} draw from among the rest of the slips in the box. Which statement below is correct?
   a) it is more likely to pick out a boy's than a girl's name this time
   b) it is less likely to pick out a boy's than a girl's name this time
   c) it is equally likely to pick out a boy's as a girl's name this time
   d) it is impossible to compare the likelihood of the two outcomes

(3) The principal picks 70 names out of the box casually without looking. The 70 names are 35 girls and 35 boys. He placed the 70 slips on the table, closes his eyes and does the 71^{st} draw among the rest of the slips in the box. Which statement below is correct?
   a) it is more likely to pick out a boy's than a girl's name this time
   b) it is less likely to pick out a boy's than a girl's name this time
   c) it is equally likely to pick out a boy's as a girl's name this time
   d) it is impossible to compare the likelihood of the two outcomes

(4) Return all the names into the box and mix thoroughly again. Now, the principal picks 70 names out of the box casually without looking. The 70 names are all boys. He placed the 70 slips on the table, closes his eyes and does the 71^{st} draw among the rest of the slips in the box. Which statement below is correct?
   a) it is more likely to pick out a boy's than a girl's name this time
   b) it is less likely to pick out a boy's than a girl's name this time
   c) it is equally likely to pick out a boy's as a girl's name this time
   d) it is impossible to compare the likelihood of the two outcomes

III5 There are 8 red marbles and 16 black marbles in bag A. There are 50 red marbles and 70 black marbles in bag B. Mix the marbles in each bag thoroughly. Close your eyes and suppose you want to pull out a black marble. Which statement below is correct?

a) the likelihood of pulling out a black marble from bag A is greater than that from bag B
b) the likelihood of pulling out a black marble from bag A is less than that from bag B
c) the likelihood of pulling out a black marble from bag A is the same as that from bag B
d) it is impossible to compare the likelihood of the two outcomes

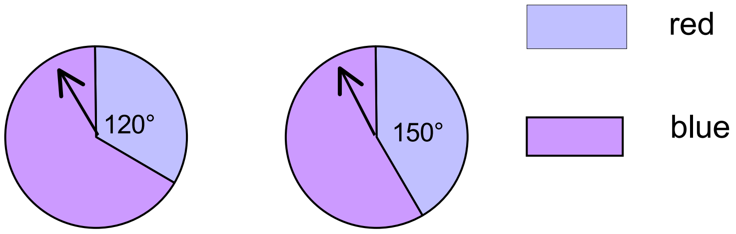
III6 There are 21 red marbles and 8 black marbles in bag A. There are 210 red marbles and 80 black marbles in bag B. Mix the marbles in each bag thoroughly. Close your eyes and suppose you want to pull out a black marble. Which statement below is correct?

a) the likelihood of pulling out a black marble from bag A is greater than that from bag B
b) the likelihood of pulling out a black marble from bag A is less than that from bag B
c) the likelihood of pulling out a black marble from bag A is the same as that from bag B
d) it is impossible to compare the likelihood of the two outcomes

III7 There are 8 red marbles and 16 black marbles in bag A. There are 500 red marbles and 100 black marbles in bag B. Mix the marbles in each bag thoroughly. Close your eyes and suppose you want to pull out a black marble. Which statement below is correct?

a) the likelihood of pulling out a black marble from bag A is greater than that from bag B
b) the likelihood of pulling out a black marble from bag A is less than that from bag B
c) the likelihood of pulling out a black marble from bag A is the same as that from bag B
d) it is impossible to compare the likelihood of the two outcomes

III8

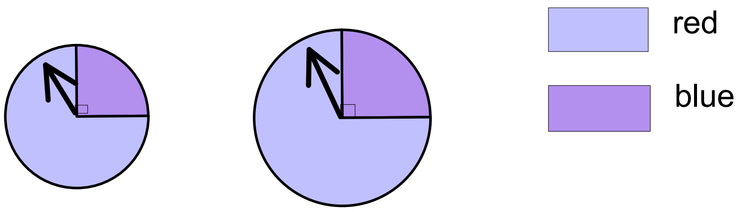


spinner A spinner B

Spin each spinner's arrowhead with all your strength. Suppose you want the arrowhead to stop in the blue part. Which statement below is correct?

a) the arrowhead of spinner A is more likely to stop in the blue part than that of spinner B
b) the arrowhead of spinner A is less likely to stop in the blue part than that of spinner B
c) the arrowhead of spinner A is equally likely to stop in the blue part as that of spinner B
d) it is impossible to compare the likelihood of the two outcomes

III9

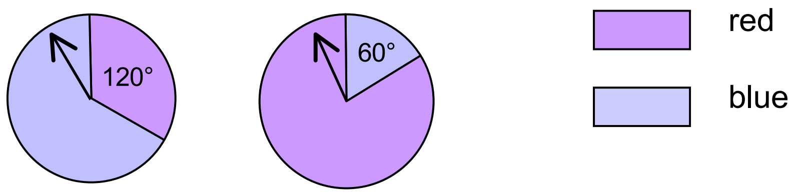


spinner A spinner B

Spin each spinner's arrowhead with all your strength. Suppose you want the arrowhead to stop in the blue part. Which statement below is correct?

a) the arrowhead of spinner A is more likely to stop in the blue part than that of spinner B
b) the arrowhead of spinner A is less likely to stop in the blue part than that of spinner B
c) the arrowhead of spinner A is equally likely to stop in the blue part as that of spinner B
d) it is impossible to compare the likelihood of the two outcomes

III10



spinner A spinner B

Spin each spinner's arrowhead with all your strength. Suppose you want the arrowhead to stop in the blue part. Which statement below is correct?

a) the arrowhead of spinner A is more likely to stop in the blue part than that of spinner B
b) the arrowhead of spinner A is less likely to stop in the blue part than that of spinner B
c) the arrowhead of spinner A is equally likely to stop in the blue part as that of spinner B
d) it is impossible to compare the likelihood of the two outcomes

Category IV items:

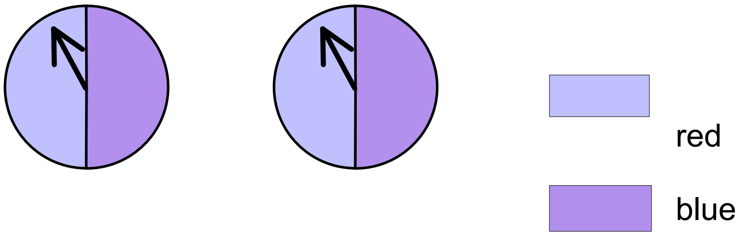
IV1 There are 2 white marbles and 2 black marbles in bag A. There are also 2 white marbles and 2 black marbles in bag B. Mix the marbles in each bag thoroughly. You put your hands in the two bags and pull out a marble from each bag without looking. Here is a possible outcome: a white marble was drawn from bag A and a white marble was drawn from bag B. Are there any other outcomes that are possible to happen? If there are, please go on listing all the outcomes possible to happen:

	Bag A	Bag B
Outcome1	white marble	white marble

IV2 There are 2 white marbles and 2 black marbles in bag A. There are also 2 white marbles and 2 black marbles in bag B. Mix the marbles in each bag thoroughly. Put your hands in the two bags and pull out a marble from each bag without looking. Which statement below is correct?

a) it is most likely that both marbles are white
b) it is most likely that both marbles are black
c) it is most likely that one marble is white and the other one is black
d) it is impossible to indicate which one is the most likely among the three outcomes

IV3

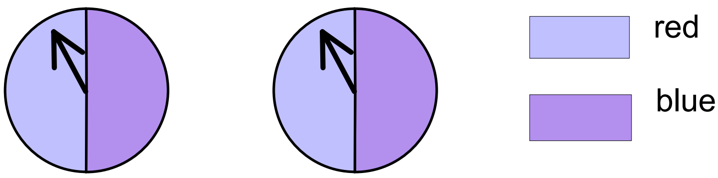


spinner A spinner B

Spin each spinner's arrowhead with all your strength. Here is a possible outcome. The arrowhead of spinner A stops in the red part and the arrowhead of spinner B also stops in the red part. If we neglect the situations that the arrowheads stop on the border, are there any other outcomes possible to happen? If there are, please go on listing the all outcomes possible to happen:

	Spinner A	Spinner B
Outcome1	stops in the red part	stops in the red part

IV4



spinner A spinner B

Spin each spinner's arrowhead once with all your strength. Which statement below is correct?

a) it is most likely that both the arrowheads stop in the red part
b) it is most likely that both the arrowheads stop in the blue part
c) it is most likely that one arrowhead stops in the red part and the other arrowhead in the blue part
d) it is impossible to indicate which one is the most likely among the above three outcomes

IV5 There are 8 red marbles and 16 black marbles in bag A. There are 50 red marbles and 70 black marbles in bag B. Mix the marbles in each bag thoroughly. Put your hands in two bags and pull out a marble from each bag without looking. Which statement below is correct?

a) it is most likely that both marbles are red
b) it is most likely that both marbles are black
c) it is most likely that one marble is red and the other one is black
d) it is impossible to indicate which one is the most likely among the three outcomes

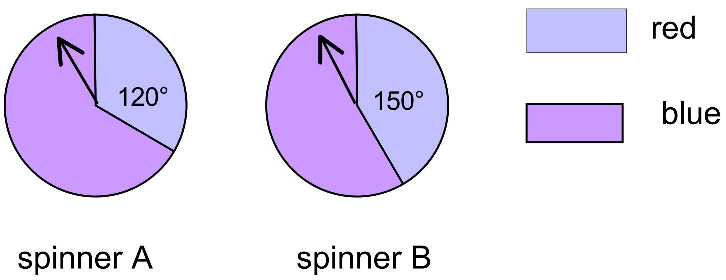
IV6 There are 21 red marbles and 8 black marbles in bag A. There are 210 red marbles and 80 black marbles in bag B. Mix the marbles in each bag thoroughly. Put your hands in two bags and pull out a marble from each bag without looking. Which statement below is correct?

a) it is most likely that both marbles are red
b) it is most likely that both marbles are black
c) it is most likely that one marble is red and the other one is black
d) it is impossible to indicate which one is the most likely among the three outcomes

IV7 There are 8 red marbles and 16 black marbles in bag A. There are 500 red marbles and 100 black marbles in bag B. Mix the marbles in each bag thoroughly. Put your hands in two bags and pull out a marble from each bag without looking. Which statement below is correct?

a) it is most likely that both marbles are red
b) it is most likely that both marbles are black
c) it is most likely that one marble is red and the other one is black
d) it is impossible to indicate which one is the most likely among the three outcomes

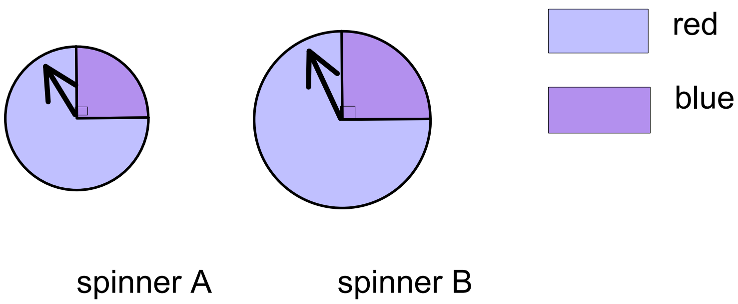
IV8



Spin each spinner's arrowhead with all your strength. Which statement below is correct?

a) it is most likely that both the arrowheads stop in the red part
b) it is most likely that both the arrowheads stop in the blue part
c) it is most likely that one arrowhead stops in the red part and the other arrowhead stops in the blue part
d) it is impossible to indicate which one is the most likely among the three outcomes

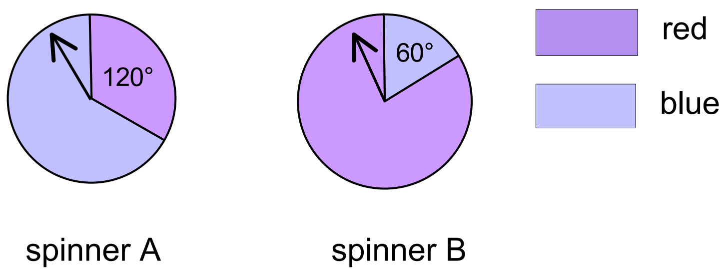
IV9



Spin each spinner's arrowhead with all your strength. Which statement below is correct?

a) it is most likely that both the arrowheads stop in the red part
b) it is most likely that both the arrowheads stop in the blue part
c) it is most likely that one arrowhead stops in the red part and the other arrowhead stops in the blue part
d) it is impossible to indicate which one is the most likely among the three outcomes

IV10



Spin each spinner's arrowhead with all your strength. Which statement below is correct?

a) it is most likely that both the arrowheads stop in the red part
b) it is most likely that both the arrowheads stop in the blue part
c) it is most likely that one arrowhead stops in the red part and the other arrowhead stops in the blue part
d) it is impossible to indicate which one is the most likely among the three outcomes

Appendix B

Pre-test Items for Teaching Intervention
(English and Chinese Versions)

The following are some general notes on the items:

1. For each item students were asked to indicate their confidence. This was presented at the end of each item in the following format. The students were asked to tick the appropriate box.

no confidence	little confidence	a lot of confidence	full confidence
☐	☐	☐	☐

2. For each item students were asked to explain their answer.

3. For items J1 and J2 the students were asked to put a tick in the appropriate box.

impossible	possible	certain
☐	☐	☐

4. Sufficient space was left to answer the questions, with a maximum of 2 items per page.

5. The questionnaires were administered in Chinese so the items were written to be as clear as possible in Chinese. In translating them back into English the final English version may appear slightly awkward due to differences between the languages, both in wording and structure. The translation was done in this manner to ensure that the meaning was not changed due to the translation.

COVER PAGE GIVEN TO STUDENTS FOR TEST

1. In the multiple-choice items if all the given choices seem wrong you are allowed to give an answer that does not appear in the choices and that you believe is right. But you must still give a reason.

2. Please give your reason to each item.

School ________________ Grade ____________ Class ___________

Name_____________ Sex ______ Date of birth______________

J1 Three six-sided normal dice are rolled once. The number on the top of each die is the number rolled. Please indicate whether the outcome, "all three numbers rolled are odd numbers" is **impossible**, **possible** or **certain** to happen. Tick where appropriate.

J2 There are three packs of cards without the picture cards and Jokers. One card is drawn from each pack of shuffled cards. The Ace is worth 1, 2 is worth 2, 3 is worth 3,…, 10 is worth 10. Please indicate whether the outcome, "all three cards drawn are smaller than 11" is **impossible**, **possible** or **certain** to happen. Tick where appropriate.

J3 A mathematician filled a bag with some black and white marbles. He didn't really know how many black marbles and white marbles were in the bag. After mixing them thoroughly he took a look and predicted that "if I pull out a marble from the bag without looking, the chance that it will happen to be white is 50%." Which of the following has the closest meaning to "the chance that it will happen to be white is 50%"?

a) the marble pulled out may be white or may be black. The mathematician doesn't really know what the result will be
b) suppose that the game is repeated 10 times, white marbles are pulled out around 5 times
c) suppose that the game is repeated 100 times, white marbles are pulled out exactly 50 times
d) suppose that the game is repeated 100 times, white marbles are pulled out around 50 times

J4 A weather forecaster said that "tomorrow there is a 30% chance that it will rain." The next day it doesn't rain. Do you think the forecaster's prediction is accurate or not?

J5 Medical research indicated that about 1 in 20 people develop a particular illness. However, in a sample of 322 people, it was found that 39 had developed this disease. Are these two pieces of information in conflict? If you really want to get a clearer picture of the real incidence of this disease, what will you do? Why?

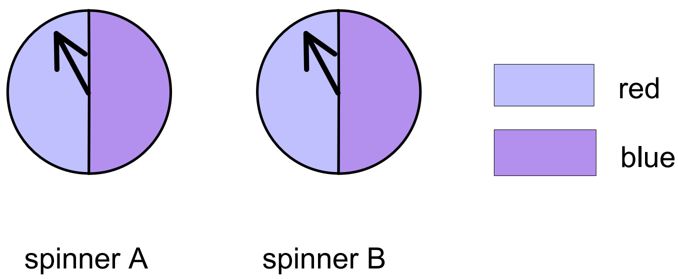
J6 A school has 500 girls and 550 boys in it. Each pupil's name is written on a piece of paper and all the names are put into a box and mixed thoroughly. The principal picks 1 name out of the box casually without looking. Which statement below is correct?

a) it is more likely to pick out a boy's than a girl's name
b) it is less likely to pick out a boy's than a girl's name
c) it is equally likely to pick out a boy's as a girl's name
d) it is impossible to compare the likelihood of the two outcomes

J7 There are 9 red marbles and 18 black marbles in bag A. There are 60 red marbles and 80 black marbles in bag B. Mix the marbles in each bag thoroughly. Close your eyes and suppose you want to pull out a black marble. Which statement below is correct?

a) the likelihood of pulling out a black marble from bag A is greater than that from bag B
b) the likelihood of pulling out a black marble from bag A is less than that from bag B
c) the likelihood of pulling out a black marble from bag A is the same as that from bag B
d) it is impossible to compare the likelihood of the two outcomes

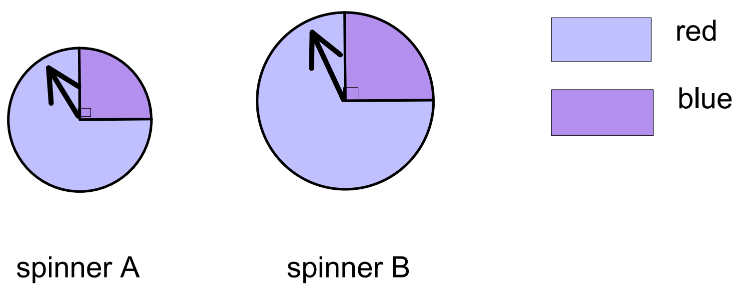
J8



Spin each spinner's arrowhead with all your strength. Which statement below is correct?

a) it is most likely that both the arrowheads stop in the red part
b) it is most likely that both the arrowheads stop in the blue part
c) it is most likely that one arrowhead stops in the red part and the other arrowhead in the blue part
d) the above three outcomes have the same likelihood
e) it is impossible to measure the likelihood of the three outcomes

J9



Spin each spinner's arrowhead with all your strength. Suppose you want the arrowhead to stop in the blue part. Which statement below is correct?

a) the arrowhead of spinner A is more likely to stop in the blue part than that of spinner B
b) the arrowhead of spinner A is less likely to stop in the blue part than that of spinner B
c) the arrowhead of spinner A is equally likely to stop in the blue part as that of spinner B
d) it is impossible to compare the likelihood of the two outcomes

Appendix C

Post-test Items for Teaching Intervention
(English and Chinese Versions)

The following are some general notes on the items:

1. For each item students were asked to indicate their confidence. This was presented at the end of each item in the following format. The students were asked to tick the appropriate box.

no confidence	little confidence	a lot of confidence	full confidence
☐	☐	☐	☐

2. For each item students were asked to explain their answer.

3. For items K1 and K2 the students were asked to put a tick in the appropriate box.

impossible	possible	certain
☐	☐	☐

4. Sufficient space was left to answer the questions, with a maximum of 2 items per page.

5. The questionnaires were administered in Chinese so the items were written to be as clear as possible in Chinese. In translating them back into English the final English version may appear slightly awkward due to differences between the languages, both in wording and structure. The translation was done in this manner to ensure that the meaning was not changed due to the translation.

COVER PAGE GIVEN TO STUDENTS FOR TEST

1. In the multiple-choice items if all the given choices seem wrong you are allowed to give an answer that does not appear in the choices and that you believe is right. But you must still give a reason.

2. Please give your reason to each item.

School ________________ Grade ____________ Class ___________

Name_____________ Sex ______ Date of birth______________

How many of the six activity classes did you miss?

K1 Three six-sided normal dice are rolled once. The number on the top of each die is the rolled number. Please indicate whether the outcome "all three numbers rolled are even numbers" is **impossible**, **possible** or **certain** to happen. Tick where appropriate.

K2 There is a pack of cards without the picture cards and Jokers. One card is drawn from the pack of shuffled cards. The Ace is worth 1, 2 is worth 2, 3 is worth 3,…, 10 is worth 10. Please indicate whether the outcome, "the card drawn is smaller than 11" is **impossible**, **possible** or **certain** to happen. Tick where appropriate.

K3 A mathematician filled a bag with some black and white marbles. He didn't really know how many black marbles and white marbles were in the bag. After mixing them thoroughly he took a look and predicted that "if I pull out a marble from the bag without looking, the chance that it will happen to be white is 50%." Which of the following has the closest meaning to "the chance that it will happen to be white is 50%"?

a) the marble pulled out may be white or may be black. The mathematician doesn't really know what the result will be
b) suppose that the game is repeated 10 times, white marbles are pulled out around 5 times
c) suppose that the game is repeated 100 times, white marbles are pulled out exactly 50 times
d) suppose that the game is repeated 100 times, white marbles are pulled out around 50 times

K4 Before the final match, the coach predicted that "based on the information I have, our chance of winning the game is 30%." The game is played and the team loses. Do you think the coach's prediction is accurate or not?

K5 Medical research indicated that about 1 in 20 people develop a particular illness. However, in a sample of 322 people, it was found that 39 had developed this disease. Are these two pieces of information in conflict? If you really want to get a clearer picture of the real incidence of this disease, what will you do? Why?

K6 A school has 400 girls and 440 boys in it. Each pupil's name is written on a piece of paper and all the names are put into a box and mixed thoroughly. The principal picks 1 name out of the box casually without looking. Which statement below is correct?

a) it is more likely to pick out a boy's than a girl's name
b) it is less likely to pick out a boy's than a girl's name
c) it is equally likely to pick out a boy's as a girl's name
d) it is impossible to compare the likelihood of the two outcomes

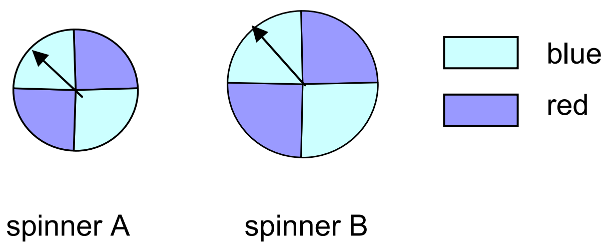
K7 There are 8 red marbles and 16 black marbles in bag A. There are 50 red marbles and 70 black marbles in bag B. Mix the marbles in each bag thoroughly. Close your eyes and suppose you want to pull out a black marble. Which statement below is correct?

a) the likelihood of pulling out a black marble from bag A is greater than that from bag B
b) the likelihood of pulling out a black marble from bag A is less than that from bag B
c) the likelihood of pulling out a black marble from bag A is the same as that from bag B
d) it is impossible to compare the likelihood of the two outcomes

K8 There are 2 white marbles and 2 black marbles in bag A. There are also 2 white marbles and 2 black marbles in bag B. Mix the marbles in each bag thoroughly. Put your hands in the two bags and pull out a marble from each bag without looking. Which statement below is correct?

a) it is most likely that both marbles are white
b) it is most likely that both marbles are black
c) it is most likely that one marble is white and the other one is black
d) the above three outcomes have the same likelihood
e) it is impossible to measure the likelihood of the three outcomes

K9



Spin each spinner's arrowhead with all your strength. Suppose you want the arrowhead to stop in the blue part. Which statement below is correct?

a) the arrowhead of spinner A is more likely to stop in the blue part than that of spinner B
b) the arrowhead of spinner A is less likely to stop in the blue part than that of spinner B
c) the arrowhead of spinner A is equally likely to stop in the blue part as that of spinner B
d) it is impossible to compare the likelihood of the two outcomes

Appendix D

Sample of the Workbook (for the first day)

Notes on the use of the workbook.

1. The workbook was given to the students at the beginning of the teaching intervention.

2. It contains the information that was the basis for the discussions and activities used that day. It was meant as an aid to the class and not as a workbook that students would go through item by item.

Part One Distinguish between the meaning of the terms "impossible", "possible" and "certain"

(1) "Impossible" means without any chance of happening at all, or the chance of happening is 0. For example, when rolling an ordinary six-sided die the outcome "the number rolled is bigger than 7" is impossible and its chance for happening is 0. That is, if you make 60000 rolls, in none of the rolls you will get a number bigger than 7.

(2) "Certain" means definitely happen, or the chance of happening is 100%. For example, to roll an ordinary six-sided die, the outcome "the number rolled is smaller than 7" is certain to happen and its chance for happening is 100%. That is, if you make 60000 rolls, in all 60000 rolls you will get a number smaller than 7.

(3) "Possible" means sometimes it happens and sometimes it doesn't, or the chance of happening is between 0 and 100%. For example, to roll an ordinary six-sided die, the outcomes:

"the number rolled is an odd number" is possible and its chance is about 30000 times in 60000 trials.

"the number rolled is 1" is also possible and its chance is about 10000 times in 60000 trials.

"the number rolled is smaller than or equals to 5" is another possible outcome and its chance is about 50000 times in 60000 trials.

All three likelihoods are located between 0 and 100%.

In order to judge whether an outcome is impossible, possible or certain to happen, you need to consider all the possible outcomes of a trial.

Exercises: Do you agree with the following sayings?

(1) An outcome with a quite high chance is certain to happen.

(2) An outcome with a quite low chance is impossible to happen.

(3) Only those outcomes with over a 90% chance are possible, lower than that percentage is impossible.

(4) "The number rolled is an odd number" is impossible, as there are some even numbers on a die as well.

(5) "The number rolled is an odd number" is certain, as there are some odd numbers on a die.

Part Two Activity: You are given an ordinary six-sided die. For each roll, the number you will roll depends on luck. However, is there any pattern that exists behind the random phenomena?

Discuss the following questions before the activity.

(1) Someone said that "6" is the most difficult number to be rolled out. Do you agree with this?

(2) Another person said that rolling a die depends wholly on luck. You may get 200 "1"s or 500 "1"s. No patterns. What's your opinion?

(3) Another person said that the result must be 100 "1"s, 100 "2"s,…, 100 "6"s, as the six numbers have the same chance. Do you think this is correct?

Now, start the activity. Support or modify your answers to the above questions with the data you collect.

(1) Make 25 rolls. How many "1"s, "2"s,…,and "6"s are rolled? Fill the six data in the column "My 1st 25 trials".

(2) Make another 25 rolls. Fill the data you collect this time in the column "2nd 25 trials".

(3) Summarise the results of your two turns and fill in the column "Summary of my 50 trials".

(4) Summarise your partner's results and your results and fill them in the column " My group's 100 trials".

(5) Summarise six groups' results and fill them in the column " 6 groups' 600 trials".

(6) Do more trials and fill the data in the appropriate columns.

All possible results	My 1st 25 trials	2nd 25 trials	Summary of my 50 trials	My group's 100 trials	6 groups' 600 trials	2200 trials	6000 trials	9999 trials
"1" is rolled	/ 25 trials	/ 25 trials	/ 50 trials	/ 100 trials	/ 600 trials			
"2" is rolled	/ 25 trials	/ 25 trials	/ 50 trials	/ 100 trials	/ 600 trials			
"3" …	/ 25 trials	/ 25 trials	/ 50 trials	/ 100 trials	/ 600 trials			
"4" …	/ 25 trials	/ 25 trials	/ 50 trials	/ 100 trials	/ 600 trials			
"5" …	/ 25 trials	/ 25 trials	/ 50 trials	/ 100 trials	/ 600 trials			
"6" …	/ 25 trials	/ 25 trials	/ 50 trials	/ 100 trials	/ 600 trials			

Answer the following questions based on the data table:

(1) Were your previous answers supported or not? Please explain.

(2) Do you think anybody can predict which number will be rolled in the first roll? In the 100th roll? In the 6000th roll?

(3) Which number(s) occurs the least/most often in your 50 trials? (See the 4th column) What's the difference in terms of percentage? For the 2200 trials, which number occurs the least/most often? What's the difference in terms of percentage? Do you find that this difference getting bigger or smaller when we do more and more trials?

(4) Now, do you know the chances of "1", "2",…,"6" being rolled?

(5) What are your conclusions based on the data that the table shows? Do you have any conjectures that you would like to share with your classmates?

Exercises: Answer the following questions based on the collected data

(1) What's the chance of the number rolled being bigger than "6"?

(2) What's the chance of the number rolled being smaller than "7"?

Problems:

1. Please list all the possible outcomes of the following experiments:

   (1) Flipping an ordinary coin

   (2) Rolling an ordinary four-sided die

   (3) Taking a card from a pack of shuffled playing cards without the picture and joker cards

2. Indicate whether the following outcomes are impossible, possible or certain to happen:

   (1) flipping two ordinary coins and getting two heads

   (2) rolling two ordinary six-sided dice and getting two "6"s

   (3) taking a card from a pack of shuffled playing cards without the picture and joker cards and the card is smaller than 8.

3. Roll an ordinary four-sided die. Indicate an outcome that is impossible to happen and another outcome that is certain to happen.