MOKAEANE V. POLAKI

DEALING WITH COMPOUND EVENTS

Chapter 8

> Instruction can lead students to actively experience the conflicts between their primary intuitive schemata and the particular types of reasoning specific to stochastic situations. (Fischbein & Schnarch, 1997, p. 104)

1. INTRODUCTION

The ability to make valid probability predictions in the context of compound events (e.g. tossing a coin and throwing a six-sided die) is a key learning goal for the middle school (National Council of Teachers of Mathematics [NCTM], 2000). More specifically, the NCTM declares that middle school students should be able to compute probabilities for simple compound events, including the expected number of occurrences of a target event for a certain number of trials. When adequately established, this knowledge base will serve as a basis for dealing with more complex compound events and other probabilistic situations including conditional and independent events in the higher grades (Tarr & Lannin, this volume).

According to Hogg and Tanis (1997), compound events are events such as A or B, A and B, and the complement of A; however the term also refers to a two-stage or two-dimensional random experiment such as the simultaneous rolling of a die and the tossing of a coin. It is this latter usage of compound events or compound random experiments, as they are sometimes called, that will be the focus of this chapter. By way of contrast, simple events or simple random experiments are one-dimensional random experiments that deal with situations such as the tossing of a single coin. Hence in compound events we are dealing with pairs of outcomes (usually ordered pairs), whereas in simple events we are dealing with single outcomes. Although the terms one-dimensional and two-dimensional are referenced in the literature (English, 1990; 1993; Zimmermann & Jones, 2002), in this chapter we will use the terms simple and compound events (see also Batanero & Sanchez; Pratt; Watson; this volume).

Graham A. Jones (ed.), Exploring probability in school: Challenges for teaching and learning. 191-214.

This chapter will analyze elementary and middle school students' ability to generate sets of outcomes associated with compound events, and will examine some research on the impact of instruction on the learning of both theoretical and experimental probability (Benson & Jones, 1999; Jones, Langrall, Thornton, & Mogill, 1999; Lecoutre, 1992; Piaget & Inhelder, 1951/1975; Pratt, 2000; Polaki, Lefoka & Jones, 2000; Polaki, 2002a; Speiser & Walter, 1998; Vidakovic, 1998; Watson & Moritz, 1998). Whereas the theoretical probability of an event is based on an analysis of sample space composition and uses symmetry, number, or simple geometric measures to determine the likelihood of an event, the experimental probability of an event has a frequentist orientation. It is based on experimentation or simulation and uses relative frequency to determine the likelihood of an event (Jones, Langrall et al., 1999). The chapter will also explore various learning experiences that might be used to nurture or support the development of students' thinking in dealing with compound events. More specifically, it will focus on understanding students' probabilistic thinking when they deal with simple and compound events in both interview and instructional settings.

2. UNDERSTANDING STUDENTS' PROBABILISITIC THINKING

As mentioned in the opening paragraph, understanding compound experiments requires one to be able to (a) generate complete sets of outcomes for each experiment, and (b) use sample space symmetry, composition or experimentation as a basis for making probability predictions. Accordingly, the concepts of sample space and probability of an event will constitute a context for exploring students' ability to deal with compound events. With these concepts in mind, reference will be made to the Probability Thinking Framework (see Figure 1, Polaki et al., 2000) which was an extension of earlier research (e.g., Jones, Langrall et al., 1997) that described students' probabilistic thinking across five constructs: sample space, probability of an event, probability comparisons, conditional probability, and independence. Validation of the Probability Thinking Framework (Framework) with Basotho students (Polaki et al., 2000) suggested the existence of four levels of probabilistic thinking: subjective (Level 1), transitional (Level 2), informal quantitative (Level 3), and numerical (Level 4). These four levels were found to be consistent with Case's (1996) more general cognitive model and in essence suggested that Case's model could be applied to probabilistic thinking in addition to the three knowledge domains previously examined by Case: numerical, spatial, and narrative thinking.

Construct	**Level 1 Subjective**	**Level 2** **Transitional**	**Level 3** **Informal quantitative**	**Level 4** **Numerical**
SAMPLE SPACE	— Lists an incomplete set of outcomes for a 1-dimensional experiment	— Lists a complete set of outcomes for a 1-dimensional experiment, and sometimes for a 2-dimensional experiment	— Consistently lists the outcomes for a 2-dimensional experiment using a partially generative strategy	— Adopts and applies a generative strategy to provide a complete listing of the outcomes for a 2-dimensional case
PROBABILITY OF AN EVENT	— Predicts most/least likely event based on subjective judgments — Provides deterministic or idiosyncratic responses	— Predicts most/least likely event based on quantitative judgement (often invalid) but may revert to subjective judgments — Acknowledges uncertainty without quantifying it.	— Predicts most/least likely events based on quantitative judgements — Uses numbers informally to compare probabilities	— Predicts most/least likely events for 1- and simple 2- dimensional experiments — Assigns a numerical probability to an event (a real probability or a form of odds)
PROBABILITY COMPARISONS	— Uses subjective judgements to compare the probabilities of an event in 2 different sample spaces. — Cannot distinguish "fair" probability situations from "unfair" ones. — Provides deterministic or idiosyncratic response	— Makes probability comparisons based on quantitative judgements-not always correct. — Cannot distinguish "fair" probability situations from "unfair" ones. — Makes part-part comparisons — Acknowledges uncertainty without quantifying it	— Uses valid quantitative reasoning to explain comparisons and invents own way of expressing the probabilities. — Uses quantitative reasoning to distinguish "fair" and "unfair" probability situations — Makes part-part and part-whole comparisons	— Assigns a numerical probability and makes a valid comparison
CONDITIONAL PROBABILITY	— Uses subjective reasoning in interpreting with and without replacement situations. — Provides deterministic or idiosyncratic responses	— Recognizes that the probability of some events change in a non-replacement situation, however recognition is incomplete and is usually restricted only to events that have previously occurred — Makes part-part comparisons — Acknowledges uncertainty without quantifying it	— Recognizes that the probabilities of all events change in a non-replacement situation — Can quantify changing probabilities in a non-replacement situation — Makes part-part and part-whole comparisons	— Assigns numerical probabilities in replacement and non-replacement situations — Uses numerical reasoning to compare the probability of events before and after each trial in replacement and non-replacement situations
INDEPENDENCE	— Has a predisposition to consider that consecutive events are always related. — Has a pervasive belief that one can control the outcome of an experiment	— Begins to recognize that consecutive events may be related or unrelated — Uses the distribution of out-comes from previous trials to predict the next outcome (representativeness) — Makes part-part comparisons — Acknowledges uncertainty without quantifying it	— Can differentiate independent events in with and without replacement situations — May revert to strategies based on representativeness — Makes part-part and part-whole comparisons — Uses notion of complement	— Uses numerical probabilities to distinguish independent and dependent events.

Figure 1. Probability Thinking Framework

In order to fully appreciate the conceptual difficulties students experience when challenged to generate sets of outcomes for compound experiments, it is important that we first examine what research says about students' ability to generate sets of outcomes for random situations involving simple events.

Sample Space: Simple Events

Some middle school students, depending on their experience with random phenomena, may have difficulty in listing the complete set of outcomes for a random experiment (Polaki et al., 2000; Green, 1983). With regard to simple experiments, research in this knowledge domain (e.g. Polaki et al., 2000) has shown that when challenged to generate the sample space, students operating at Level 1 (Figure 1) typically provide incomplete sets of outcomes and often justify their responses subjectively. Figure 2 shows examples of assessment tasks associated with simple events.

---

1. Show a spinner with 4 evenly-spaced colors: Red, Green, Blue, and Yellow.
   a) You got ....on the first spin, what colors could you get if you spin again?
   b) Write them down.
   c) Can you explain that to me?
   d) Which color are you most likely to get if you spin again?
2. Show a six-sided fair die, and allow the student to roll the die.
   a) What did you get?
   b) Write down all the outcomes you could get if you roll the die again.
   c) Can you explain that to me?
   d) How would you describe the chance of obtaining an even number when you roll the die?

Figure 2. Examples of Tasks Based on Simple Events

Polaki et al. (2000) found that when the students exhibiting Level 1 thinking were asked to list the set of outcomes for a spinner with four colors, they were more inclined to give one outcome, arguing that the spinner would land on their favorite color. When these students were challenged to list all possible colors to which the pointer could land after landing on one of four colors on the first spin (see Figure 2: Item 1), Polaki et al. observed that the students excluded the color they got on the first spin, arguing that they could not get it again since they got it on the first spin. Furthermore, when asked to predict the most likely color (Item 1 (d), Figure 2), these students picked one of the 4 colors, often arguing that the one they had chosen was their favorite.

Similarly, in response to Item 2b (Figure 2), where the students were asked to list all possible outcomes on the second roll of a fair die, Level 1 students were more inclined to only mention 5 of the six equally-likely outcomes, arguing that they could no longer get the outcome they obtained on the first trial. Jones, Langrall et al. (1999) termed this "the sample space misconception." Furthermore, these researchers observed that this thinking tendency remained resilient even when subjected to a carefully designed instructional program (see Langrall & Mooney, this volume). Apart from this misconception, students tended to give a wide range of subjective reasons including the claim that their incomplete lists of outcomes were based on personal preference or that they knew for sure that things would happen the way they had predicted (deterministic perspective). Similar findings were recorded in Jones et al. (1997). Unlike their Level 1 (subjective) counterparts, students exhibiting Level 2 (transitional) thinking were often able to provide complete sets of outcomes for simple random events like those incorporated into Items 1 and 2 (Figure 2). It is important to note that students at Level 3 (informal quantitative) and Level 4 (quantitative) experienced no difficulty in listing complete sets of outcomes for simple random experiments. However, students operating at the Levels 1 through 4 experienced varying degrees of difficulty when challenged to provide complete sets of outcomes for compound random events.

Two possible interpretations have been documented to explain the observed features of students' thinking with regard to generating sets of outcomes for simple experiments. Jones et al. (1997) claimed that the tendency of Level 1 (subjective) students to provide an incomplete set of outcomes with subjective justifications was consistent with Biggs & Collis' (1991) prestructural thinking in that the students seemed to be distracted by an irrelevant aspect: generally their preoccupation with their favorite color. Using a different cognitive lens, Polaki et al. (2000) argued that this type of thinking corresponded to Case's pre-dimensional thinking level in the sense that the absence of a mental counting line had made it impossible for students at this level to construct a part-part schema (conceptual structure that enables the learner to compare or order parts) that would enable them to coordinate the organization and numbering of the elements of the sample space. Whereas a part-part schema is a conceptual structure that enables the learner to compare or order parts of a whole, a part-whole schema is a related structure that makes it possible for the learner to compare parts to a whole (Lamon, 1999). For this reason, such students provided an incomplete sample space, supporting their responses with subjective reasons. Polaki et al. also noted that the observed subjective reasoning was a result of failure to coordinate order and numbering. As further confirmation of their position, they asserted that the mechanisms successfully used in sample space tasks by students operating at Levels 2 through 4 did reveal the presence of a mental counting line and a part-part schema that enabled the students to organize and list complete sets of outcomes.

Sample Space: Compound Events

Generating complete sets of outcomes for compound random experiments presents students operating at Levels 1 through 4 of the Framework (Figure 1) with varying degrees of challenge. Figure 3 shows examples of the kinds of tasks that were used to assess thinking with respect to compound experiments. Whereas Level 1 students typically experience no success at generating sets of outcomes for compound experiments, their Level 2 counterparts show some success on this task, but often provide incomplete sets of outcomes. Polaki (2002b) concluded that whereas the existence of a mental counting line enables students operating at Level 2 to organize and generate the sample space for simple experiments, it does not enable them to generate sets of outcomes for compound events. Such events require, as will be shown shortly, the use of more than one mental counting line to organize and generate complete sets of outcomes.

Research in this knowledge domain (English, 1990, 1993; Benson & Jones, 1999; Polaki, 2002a) has further shown that, at the lowest level of sophistication, students provide incomplete sets of outcomes for compound experiments often on the basis of some subjective reasoning (e.g. personal preference) or trial-and-error strategies. For example, when challenged to list all possible ways of choosing a pair of pants and a pair of shoes from 3 pairs of pants and 2 pairs of shoes (Figure 3: Item 3), many students first try to match shoes and pants without following any systematic strategy. The final step is to go through this rather haphazard list to identify repeated pairs and delete them. Polaki (2002a) observed that this inefficient strategy might work for the less complex compound situations such as Item 3 (Figure 3) but not for the more difficult situations such as Item 4 (Figure 3).

Level 3 (informal quantitative) students differed from their Level 1(subjective) and Level 2 (transitional) counterparts in that they were able to provide complete sets of outcomes for compound experiments using a partially generative strategy (Polaki et al., 2000; Jones et al., 1997). Whereas the Level 2 students would generate an incomplete set of outcomes for tossing a fair coin and a fair die (Figure 3: Item 1) without following any order or strategy, Level 3 students would first produce 6 outcomes by a kind of alternating of a H and a T; each time picking up numbers they had not picked up from the 6 possible outcomes of a die as follows: "H 1, T 2, T 1, H

1. Show a fair coin and a six-sided fair die, allow the student to roll them at the same time.
a) What did we get? Write down all the outcomes you could get when you roll the die and flip the coin at the same time.
b) Can you explain that to me?
c) What is the probability of obtaining an H and an even number?
2. You and I are playing a game. You toss a fair coin and win a point every time it turns up heads. I roll a six-sided fair die and win a point every time I get an even number (2, 4, or 6). If you wanted to win the game, would you choose a coin or a die? How did you decide?
3. Thabo has 3 different pairs of pants: 1 grey, 1 khaki, and 1 white. Furthermore, he has two pairs of shoes: 1 black and 1 brown. Suppose he chooses a pair of pants and a pair of shoes without looking:
(a) How many possible combinations of pants and shoes can Thabo choose to wear?
(b) List all possible combinations of pants and shoes that Thabo can choose to wear.
4. Palesa plans to eat lunch at Speak Easy Restaurant. Three types of Lunch are available: fish (with potato chips), chicken (with rice), and beef stew (with papa). Each lunch is served with 1 of the following beverages: coke, pepsi, fanta, monis, sprite, appletizer, and grapetizer.
a) In how many different ways can she choose to eat at Speak Easy?
b) List all possible ways in which Palesa can choose to eat at Speak Easy.

Figure 3. Sample Tasks Based on Compounds Experiments

2, H 3, T 4." They would then continue in this way until they had produced a complete set of 12 possible outcomes. In contrast to the students using this partially generative approach (Level 3), those exhibiting Level 4 used a generative approach to list complete sets of outcomes for compound experiments (Polaki et al., 2000; Jones et al., 1997). For example, when asked to list all possible ways of choosing to eat at a restaurant, given 3 types of lunch and 7 different beverages (Item 4, Figure 3), these students took each type of lunch and then systematically matched it with each of the 7 beverages to produce a complete set of 21 outcomes. English (1993) termed this approach the *odometer strategy* for, like the roll over an odometer, it entails taking each possible outcome in one set (in this case types of lunch) and systematically matching it with each of the outcomes in the second set (see English, this volume, for a more detailed description). Implicit in this strategy, according to Polaki (2002b), is the ability of the learner to use a crude form of the multiplication rule in order to figure out when the set of possible outcomes is indeed complete. Clearly, this is a highly sophisticated strategy compared to the trial-and-error strategy used by students operating at Level 2.

It is appropriate to consider a possible interpretation of the partially generative strategy exhibited by students operating at Level 3 (informal quantitative). According to Polaki et al. (2000) the thinking of these students appeared to be consistent with Case's (1996) bidimensional thinking in that they were able to employ more than one counting line to do the arithmetic necessary for generating complete sets of outcomes for compound experiments. In the example described in the foregoing paragraph, the process of listing a complete set of outcomes for rolling a six-sided die and tossing a fair coin (Item 1, Figure 3) entails recognizing and counting elements of each of the sample spaces associated with the coin and the die, before integrating these into a whole. Although Jones et al. (1997) used a different psychological perspective, their interpretation is similar. They argued, in accord with Biggs & Collis (1991), that students at Level 3 were often able to focus on more than one aspect of a situation, that is, exhibit multi-structural thinking.

To interpret Level 4 students' ability to produce a complete set of outcomes for compound experiments using a generative strategy, Polaki et al. (2000) have argued that Level 4 students' thinking is more consistent with what Case (1996) described as integrated bidimensional thinking. According to Case, students using integrated bidimensional thinking are able to use and systematically coordinate arithmetical thinking using multiple counting lines. It is this *coordination* of multiple counting lines in using the odometer strategy that distinguishes Level 4 from Level 3 students. In a similar way, Jones et al (1997) argued that Level 4 thinking is more consistent with Biggs and Collis' (1991) relational thinking in the sense that students at this level are able to integrate more than one aspect of a situation into a meaningful structure. In the case of compound situations such as the tossing of two fair coins, this process entails simultaneously counting and ordering the elements of the two sets, and integrating them so as to generate a compound sample space {(H H), (H T), (T,H), (TT)}. The added use of a rough multiplication rule to figure out that the maximum number of possible outcomes is 4, confirms for these students that the sample space is indeed complete.

Probability of an Event: Simple Events

To make likelihood predictions in the context of simple events, Level 1 (subjective) students typically provided subjective responses, including idiosyncratic and deterministic reasoning (Polaki et al, 2000; Jones et al., 1997). For example, they would argue that a tail is more likely when a fair coin is tossed because it is a favorite outcome or because "it often comes up for them." In contrast, Level 2 (transitional) students showed more success at making probability predictions for simple events, and had started to use informal but valid quantitative judgments to predict the most-likely or least-likely event, albeit inconsistently. For example, when asked to describe the chance of obtaining an even number when a six-sided fair die is rolled, these students sometimes used the phrase "3 out of 6". However, they typically used this informal quantitative language rather inconsistently and at times reverted to subjective reasoning. In particular, Polaki et al. (2000) found that what Watson, Collis & Moritz (1997) termed "acknowledgment of uncertainty without quantifying it" was prevalent amongst the sample of elementary and middle school students they assessed. For instance, when asked to predict whether a head or tail was more likely to occur when tossing a fair coin, students showing this type of thinking insisted that they did not know because anything could happen.

Polaki et al. (2000) argued that whereas Level 1 students had not as yet constructed a mental counting line, their Level 2 counterparts had constructed this counting line, and it enabled them to construct a part-part schema. This structure appeared to have made it possible for Level 2 students to coordinate the notions of number and ordering needed for comparing probabilities of simple events. Thus the thinking shown by Level 1 students is more consistent with Case's prestructural thinking. In contrast, Level 2 students' ability to list complete sets of outcomes and their limited success at predicting likelihood was more consistent with Case's (1996) unidimensional thinking. In essence, the presence of a mental counting line enabled Level 2 students to construct a part-part schema that made it possible for them to coordinate number and ordering in the case of simple events but not, as we will see, for compound events that required part-part and part-whole comparisons. Whereas such skills would suffice for simple experiments (structurally more simple), they apparently did not work for compound random situations where the learner needed to think of more than one aspect of a situation and then simultaneously integrate this into a single structure (more complex). Using a different cognitive perspective, when interpreting similar findings, Jones et al. (1997) posited that the type of thinking shown by Level 2 students corresponds to Biggs and Collis' (1991) unistructural thinking in the sense that the students appeared to have engaged the task in a relevant way even though they focused on a single aspect. This explains why they can make valid probability predictions for simple events but not for compound events, which require them to coordinate more than one aspect of the random situation.

Students operating at Level 3 and 4 experienced no difficulty in making predictions in the case of simple random experiments. Furthermore, they showed greater consistency when using valid quantitative judgements to predict the most and least likely events. However, they often stopped short of using precise numerical measures (fractions) when challenged to do so. In the sense of Case (1996) the students operating at these higher levels had constructed a mental counting line that enabled them to coordinate number and the ordering of probabilities for simple events. I examine in the next section compound random experiments; these caused varying difficulties for students at all levels.

Probability of an Event: Compound Events

As we noted previously, students operating at Level 1 and Level 2 experienced great difficulty with sample space; that is in listing complete sets of outcomes for a compound random experiment. Since valid probability predictions are derived from an analysis of sample space composition or symmetry, it is logical to expect these students to struggle when challenged to make probability predictions for compound experiments. This probably explains why most of them tend to make arbitrary predictions in the context of compound events, justifying them subjectively.

By way of contrast, Level 3 and Level 4 students experienced greater success at listing complete sets of outcomes for compound events, but were very erratic when asked to make predictions of likelihood. In the sense of Case (1996), students operating at these levels (3 and 4) had, in the same way as their Level 2 counterparts, constructed a mental counting line that enabled them to order probabilities for simple events but not for compound events. These latter events require the construction and use of multiple counting lines to perform the arithmetic associated with probabilistic thinking.

In this section I have provided background on the kinds of thinking upper elementary and middle school students might bring to the classroom with regard to the generation of sample space and the making of likelihood predictions for simple and compound events. Additionally, I have attempted to provide a psychological interpretation of students' thinking, and have pointed to connections between quantitative thinking and probabilistic thinking as it pertains to both simple and compound events. The presence of mental counting lines and the subsequent construction of part-part and part-whole schemata appear to drive the development of students' probabilistic thinking.

The next section examines attempts to support or nurture the development of students' ability to deal with simple and compound random experiments in an instructional setting (e.g. Jones, Langrall et al., 1999). My intention is to describe key changes in the development of students' probabilistic thinking, and to identify some instructional strategies that might be used to develop and encourage more sophisticated forms of probabilistic thinking.

3. STUDENTS' THINKING IN AN INSTRUCTIONAL SETTING

A number of attempts have been made to study the development of students' ability to generate complete sets of outcomes and to make valid probability predictions for simple and compound random experiments (e.g. Polaki, 2002a; Jones, Thornton, Langrall & Tarr, 1999). This section provides a discussion of the observations made in an instructional program that was aimed at documenting and interpreting how upper elementary school students acquire increasingly sophisticated ideas in dealing with simple and compound random situations (Polaki 2002b). It is hoped that this discussion provides a useful picture of the type of thinking that students will bring to instruction in the middle school years. As in the previous section, notions of sample space and probability of an event are used as the context for examining the development of students' understanding in relation to simple and compound events.

Instructional Program

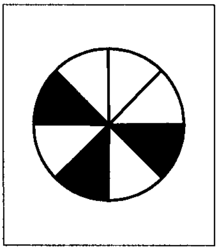
The instructional program was premised on the *cognitively guided instruction* model: according to this model research-based descriptions of students' thinking in a knowledge domain are used to inform instructional decisions (Carpenter & Fennema, 1988). Accordingly, the Framework (Figure 1) was used as the research base on students' probabilistic thinking to inform instruction that focused on simple and compound random experiments. In the sense of Jones, Thornton et al. (1999), the Framework was used in three ways: (a) planning the instructional session in that it constituted a basis for selecting and developing appropriate learning activities, (b) implementing the instructional session in that it provided a context for interpreting and classifying students' responses and interactions during instruction, and (c) assessing and monitoring students' thinking at various stages of instruction.

An overarching goal of the instructional research was to develop a detailed account of key episodes and conditions that are crucial to enabling the students to make conceptual progress in thinking about simple and compound random experiments. The intent was to formulate a learning trajectory (Cobb, 2000; Simon, 1995) that described students' expected thinking as they generated sets of outcomes and made predictions for tasks involving simple and compound events. An instructional sequence design that proved useful in evoking key changes in students' thinking entailed five phases: (a) exposing students to a game-like random situation and asking them to list possible outcomes and to make initial probability predictions; (b) asking the students to act out (play) the game-like situation a limited number of times (say 50), and then asking them to reflect on the predictions they made in the first phase; (c) simulating the game, displaying the data, discussing the results, and then asking the students to reexamine the responses they provided in the first phase; (d) examining sample composition or symmetry in order to reconsider the questions posed in the initial phase; and (e) reconciling the observations made after the computer simulation phase with the results of the analysis of sample space composition. Figure 4 shows an abridged version of the instructional session built around a simple random experiment. Figure 5 shows examples of tasks used in the instructional setting.

Sample Space: Simple Events

The discussion in this section will focus on how students' probabilistic thinking evolved from subjective reasoning to an ability to generate complete sets of outcomes for simple random experiments. The pretest indicated that all the 12 students who took part in the instructional program operated at Level 1 (subjective) thinking prior to the start of the instructional program. The reader should note that, unlike the assessment items, instructional activities did not explicitly require students to generate sets of outcomes for simple experiments. Instead, the need to focus on the set of all possible outcomes for simple experiments was implicitly called for when the students were challenged to make probability predictions. The instructor's probing questions together with discussions in whole- and small-group settings seemed to be crucial in enabling the students to construct explicit links between probability predictions and sample space composition. It was observed that at the end of the 6-week instructional period 8 (67%) students were operating at Level 4 (numerical) with regard to ability to generate sets of outcomes for simple and compound random experiments. Given that sample space and probability of an event are interrelated, the instructional session that challenged the students to make predictions on the basis of

Which Color Now?



Materials: Spinner as shown.

1. Making Conjectures
Each player chooses either a black or white color. Players take turns to spin the spinner. Each player wins a point each time the pointer lands on which chosen color. The winner, player with the greatest number of points, wins a Walkman. If you wanted to win a Walkman, which color would you choose? How did you decide? [Challenging students to make conjectures]
2. Playing the Game
Students work in groups. Two students take turns to spin the spinner. The third student records the number of times the spinner landed on black or on white. The game takes 50 trials.
3. Analyzing Sample Space Composition
Did the game turn out as you expected? Why or why not? Which color was best for winning the Walkman? Why? If we wanted to predict the winning color before playing the game, what would you suggest we do? Explain.
4. Making Extensions
Was this a fair game? Why or why not? If you think the game was unfair describe how you would design a fair game. If you think the game was fair, describe how you would design an unfair game.

Figure 4. Typical instructional session built around a Simple Random Event

sample space composition forced the students to focus more closely on the need to generate complete sets of outcomes.

Case study analyses revealed that although the majority of students made substantial progress in dealing with sample space, two students named Mpho and Tau showed a persistent belief that the outcomes of a random experiment were dependent on previous outcomes (Jones, Langrall et al., 1999). This occurred despite learning experiences that were designed to challenge this misconception. Eventually the thinking of the rest of the students progressed beyond Level 1 with respect to the listing of outcomes of simple random experiments. However, the generation of a complete set of outcomes in the case of compound experiments produced new challenges and interesting developmental patterns.

Sample Space: Compound Events

Whereas most students had little difficulty in generating complete sets of outcomes for simple random experiments, they struggled when challenged to generate sets of outcomes for compound random experiments. In order to evoke growth in students' ability to generate complete sets of outcomes, they were first asked to figure out the number of ways in which a child named Thabo could choose to wear a pair of pants and a pair of shoes: given 3 pairs of pants in the top drawer and 2 pairs of shoes in the bottom drawer (Item 3, Figure 3). Another child called Tefo was the only one who gave a correct response using a nongenerative strategy. This strategy entailed matching a pair of pants to a pair of shoes without following any system, and then checking to see if any of the items had been omitted. The rest of the students also employed this strategy, albeit without success. It is also worth noting that compared to other items involving compound events (e.g. Item 4; Figure 3), this item was a structurally easier problem as it involved two sets with small numbers of elements; namely two and three respectively.

Instead of showing the students how to do the problem, the instructor asked the students to solve a similar but more complex problem. The problem challenged students to figure out the number of ways of eating at a restaurant given 3 types of lunch and 7 types of beverages (Item 4, Figure 3). This time the instructor insisted on the need to figure out a systematic strategy for listing all possible outcomes. In so doing, the instructor established conditions that produced a cognitive conflict in the hope that this conflict would motivate the students to figure out a systematic way of listing all possible outcomes. Three types of responses came to the fore. First, the majority still used the trial-and-error strategy, and consequently gave up after listing an incomplete set of outcomes. Second, one student called Mampe came up with 18 possible outcomes after matching each type of lunch with all but one of the 7 beverages. Third, another student called Lineo followed a similar approach but used all beverages, and correctly came up with 21 possible outcomes. Mampe and Lineo had used the *odometer* strategy (English, 1993) to a different degree of accuracy. Interestingly, Lineo had not been able to use the same strategy earlier to solve a much simpler problem (Item 3, Figure 3). Additionally, both Mampe and Lineo were able to figure out the number of possible outcomes, apparently using the multiplication rule. That is, when Lineo was asked to explain how she figured out that the number of possible ways of eating at the restaurant was 21 (Item 4, Figure 3), Lineo explained: "Well, I figured out how many drinks there were.....So I multiplied 7 by 3 to get 21".

As the lesson continued, Mampe, Lineo, and Lebo were asked to present and defend their solution strategies in a whole-class discussion setting.

Whereas Mampe and Lineo had used the odometer strategy (English, 1993) to varying degrees of success, Lebo had attempted in vain to use a trial-and-error strategy to solve the problem. This discussion enabled those students who had employed trial-and-error approaches to become aware of the limitations of their approach. More importantly, it enabled the rest of the class to become aware of the odometer strategy. Indeed, subsequent lessons did suggest that the majority had begun to use the odometer strategy successfully to solve similar problems. Thus the creation of a cognitive conflict by way of posing a more challenging task when the students were experiencing difficulties with the less challenging task made it possible for them to attain a conceptual breakthrough. Additionally, discussions of these type of tasks in small- and whole-group settings helped to move the majority of the students closer to using the kinds of multiple operations that were needed for sample space tasks involving compound events.

In essence the students noticed that in listing sets of outcomes for compound events, there was no reasonable alternative to using a systematic strategy. For Item 4 (Figure 3), the arithmetic seemed to entail not only counting the number of elements in both sets (types of lunch and types of beverages) but it also entailed clever counting via the use of the multiplication rule: one needs to multiply the number of beverages (7) by the number of types of lunch to figure out the number of possible outcomes. It is important to mention that subsequent assessments indicated that 8 of the 12 students who participated in the study (67%) were indeed operating at the highest level of thinking with respect to sample space (Level 4). This was evidenced by their proficiency in listing complete sets of outcomes for compound random experiments using a generative strategy.

Probability of an Event: Simple Events

In the first phase of the instructional session focusing on simple random experiments (see Figure 4), the majority of the students claimed that the player with a black color would win for a wide range of subjective reasons, including the fact that black was a favorite color. During the second phase of this instructional program when the students were asked to play the game in small-group settings, the player with the white color happened to win in all the groups. This development coupled with discussions on how to figure out a winning strategy by examining the composition of the sample space was crucial in shaking the students' subjective stance. Although some students argued that the white color was best for winning because they lost when they chose black, their thinking changed considerably after focusing on the analysis of sample space composition. Many began to argue along these lines: "We would get 3 blacks and 5 whites. We said we look for the one that was bigger. I will say 5 out of 8." They had apparently begun to establish implicit connections between the winning color and the composition of sample space.

Although students' use of informal quantitative expressions such as "5 out of 8" to describe the chance that the pointer would land on white could lead one to believe that the students were using part-whole comparisons, it became clear that they were merely comparing parts of a whole (part-part comparison) rather than making part-whole comparisons. This came to the fore when students were challenged to deal with probability situations that definitively call for the use of a part-whole schema to order probabilities. For example, in the process of deciding whether it would be best to choose a die or a coin in Item 2 (Figure 3), most students were able to describe the probability of getting a head on a fair coin as "1 out of 2" and the probability of obtaining an even number on a die as "3 out of 6". However, some argued that they would choose a coin in order to win the game because, as they explained it, "It did not take that long to get a head". Others asserted that they would choose a die because one could get an even number many times on a die. Clearly, none of the students seemed to attach a quantitative meaning that went beyond comparing parts of different wholes (sets of outcomes for a coin and a die). It would seem, therefore, that the presence of a mental counting line (Case 1996) made it possible for the students to construct a part-part schema that enabled them to make informal quantitative comparisons in certain contexts, but not for those that clearly required part-whole comparisons.

Probability of an Event: Compound Events

The students in Polaki's (2002b) study seemed to struggle conceptually in making valid probability predictions for compound random experiments. In order to illustrate the extent of the complexity of making likelihood predictions in these situations, it is useful to look at another typical instructional experience slightly different from the one shown in Figure 4. Whereas Figure 4 describes an instructional session built around a game-like simple random event, Figure 6 summarizes key features of a game-like situation incorporating compound experiments. Both instructional sessions consisted of four learning phases. They were aimed at enabling the students to (a) base probability predictions on sample space composition, and (b) establish conceptual connections between theoretical and experimental probability.

1. Show spinner with 8 identical sectors: 3 painted black, and 5 painted white. Each player chooses either a black or white color. Players take turns to spin the spinner. Each player wins a point each time the pointer lands on a chosen color. The winner, player with the greatest number of points, wins a Walkman. If you wanted to win the game, which color would you choose? How did you decide? Was the game fair? Why or why not?
2. Materials: 2 fair coins. [Spin one coin and allow the student to spin the other]. What did we get? Write down all the outcomes you could get when you spin both coins again. Can you explain this to me? Are you more likely to get 2 heads, 2 tails, one of each, or is it the same chance? How did you decide? How would you use numbers to explain this to your friend? I did it 100 times this morning; how many times would you expect I got (a) 2 heads, (b) 2 tails, and (c) one of each? Please explain your answer.
3. Each of the two identical containers A and B has 7 bears: 2 red, 2 yellow, and 3 green. You and a friend take turns picking a bear from the two containers without looking. If there is a color match, that player wins a point. If there is no-match, the other player wins a point. If you wanted to win the game, would you choose to aim for a color match or a color mis-match? How did you decide?
4. Each of the 11 jockeys chooses one of the horses numbered: 2,3,4,5,6,7,8,9,10,11, and 12 to compete in a race from Mahlanyeng to Mafefoane. The jockeys take turns to roll two six-sided dice. Each jockey moves his horse one-step whenever the sum matches the number on his horse. The winning horse is the one that gets to Mafefoane first. If you wanted to win the race, which horse would you choose? Why?

Figure 5. Examples of tasks used in an instructional setting

The researcher began by simulating, with Minitab software, each compound random experiment for the following sequence of trials using Minitab statistical software: 10, 20, 100, 500, 1000, 5000, 10 000. Then the results of each sequence of trials were displayed in the form of pie charts and bar graphs using the Excel software. Following a full explanation of the process of a simulation, the students were shown the data display for each compound experiment. Finally the students were challenged to comment on the data, and to make connections between the probabilities derived from analysis of sample space composition or symmetry and those derived from experimentation or simulation.

1. Making Conjectures. Mpho and Teboho are playing a game, each tossing a fair coin. If there is a match (2 heads or 2 tails), Mpho wins a point. If there is no match, Teboho wins a point. Mpho says the game is unfair. Teboho disagrees, arguing that the game is fair because each of the players has a chance to win a point. What do you think? Why?
2. Playing the Game. The students work in groups of three. One chooses a match and the other chooses a no-match. The two students take turns tossing the coin. The third students record the results. Did the game turn out as you expected? Explain.
3. Examining Data Generated from a Simulation. The researcher shares the results of doing the experiment 20; 50; 100; 500; 1,000; 10,000 times using the computer. What can you say about the results? Do you think the game was fair? Why or why not?
4. Making Extensions. List all possible outcomes of this game. What was the probability of getting a match? What was the probability of getting a no-match? If you wanted to figure out whether the game was unfair or fair prior to playing the game, what would you do? Are there any connections between your answers and the data obtained from a computer?

Figure 6. Typical instructional session built around a compound random situation

In spite of the fact that most students had learned how to generate a complete sample space for compound random experiments using the odometer strategy, it became apparent that the students were not basing their probability conjectures (Item 1, Figure 6) on an analysis of sample space composition as readily as they had done in the case of simple random situations. On the contrary, the majority gave a range of incorrect responses, including the claim that the game was fair because each of the players had a chance to play; that is a chance to try winning a point. As expected, small-sample experimentation (playing the game 50 times) proved not to be very helpful in this task, because the player with a match got more points in some groups and fewer points in others, and so the results were inconclusive. In addition, long-term simulation had no real effect on students' understanding of probability because they failed to make any useful connections between theoretical probability and experimental probability. It was in the modeling of the game (focusing on an analysis of sample space composition) where students appeared to make explicit connections between probability of an event and sample space composition. They listed a complete set of outcomes for the problem described in Figure 6, namely, {(HH), (HT), (TH), (TT)}, and were eventually able to conclude correctly that the game is fair because one can have no-match "2 out of 4 times" and a match "2 out of 4 times".

In looking at the data generated from a computer simulation of the experiment described in the third phase of the instructional program (Item 3,

Figure 6), students' attention was drawn to the fact that, as the number of trials increased, the number of target outcomes (matches in this item) tended to stabilize around a fixed number (0.5). Whereas some students did recognize this observation, none of them were able to make explicit connections between this ratio and the theoretical probability of obtaining a match when two fair coins are tossed. In other words, the students could not establish connections between experimental probability and theoretical probability. This difficulty was also observed after dealing with similar game-like situations (e.g. Item 4, Figure 5). It seems that recognizing that experimental probability (relative frequency) approaches theoretical, probability (law of large numbers) as the number of trials increases was beyond the thinking of the students who took part in the instructional session. In essence, making connections between experimental and theoretical probabilities entails seeing the relationship as a limiting process; this was a highly abstract idea for these elementary school students many of whom had exhibited deficiencies in quantitative thinking and had had little experience in experimentation (see Pratt; Stohl; this volume).

Indeed research on students' understanding of simulation (e.g. Zimmerman & Jones, 2002) has shown that simulations continue to pose a lot of conceptual difficulties even for older students. Zimmerman and Jones challenged high school students to (a) assess the appropriateness of a simulation given a compound random experiment, and (b) design an adequate simulation given a compound random experiment. Results showed that the students had difficulty in responding to both tasks. More specifically, they tended to construct a simulation that would be appropriate for a single random experiment rather than for the given compound random experiment. Whereas some students cherished useful beliefs such as the fact that assumptions are necessary in making simulations, and experimental probability would approach theoretical probability as the number of trials increased, others held problematic beliefs such as simulations cannot be used to model a real world problem.

Returning to Polaki (2002b), it is apparent that despite students' consistent ability to list complete sets of outcomes for compound random experiments, and their exposure to large-sample simulations, they experience great difficulty in making predictions about compound experiments. This became even more apparent when they responded to assessment items. The following episode illustrates the nature of this difficulty. At the end of the instructional period, Thabo successfully listed a complete set of outcomes for tossing two fair coins simultaneously (Item 2, Figure 5). When asked how many times he expected to get (a) 2 heads, (b) 2 tails, and (c) one of each if the experiment was done 100 times, he said he would get 2 heads 25 times. He explained as follows: "When you multiply 25 by 4 you will get a 100." Furthermore, when challenged to use numbers to explain the probability of getting two heads he said, "100 over 25......I mean 25 over a hundred." He provided similar, though invalid, responses when challenged to predict the number of times he would expect to get "one of each" (a head-tail combination) if the experiment was done 100 times. Surprisingly, he was unable to provide similar responses for other compound random situations. It seems that Thabo was able to establish some connection between sample space and probability measures, albeit inconsistently.

The only student who made real progress in making predictions for compound processes was Lineo. Before the start of instructional activities, Lineo stated, in response to Item 2 (Figure 5), that she expected to get two tails 25 times if the experiment were done 100 times. She elaborated on this by saying," 50 is left for the head and tail, and tail and head." However, she later said that she would get "one of each" 25 times. In this latter case it was not clear whether she meant a head followed by a tail, a tail followed by a head, or a combination of the two. She had difficulty explaining her response but it seems that she had reverted to just a head and a tail (or a tail and a head) and hence was not in contradiction with her earlier assertion that "50 is left for the head and tail, and tail and head."

In another situation (Item 3, Figure 5) where Lineo was asked to decide whether it would be wise to choose a match or no-match, given 2 identical containers each with 7 bears (2 red, 2 yellow, and 3 green), Lineo correctly chose a no-match arguing that one could get a no-match many times. Apparently she was, by this time, able to mentally generate and visualize a complete set of outcomes for a compound experiment before making a valid decision. Furthermore, in a similar compound random experiment, where she was shown two identical spinners each divided into 2 equal sectors labeled 10 and 4 respectively, Lineo used a generative strategy to list all possible sums when both spinners are spun at the same time. She argued that the most likely sum was 14, and she justified this voluntarily using precise numerical measures (fractions). Her explanation was as follows: " It is 2 over 4 because it can be 10 with 4 and 4 with 10 and so you can get 14 in 2 different ways."

My research (Polaki, 2002b) suggested that the progress made by Lineo occurred because of the development of a more stable part-part schema. This schema made it possible for her to count and order elements of the sample space before making predictions. Pedagogically, the instructors' insistence on having the students focus on the composition of the sample space appeared to be critical in helping Lineo build this part-part schema. More specifically, when the students failed to recognize that Item 2 (Figure 3) represented a fair game, the instructor drew 2 identical chocolate bars one divided into 2 equal parts and the other into 6 equal parts. He then asked the students to decide which of the 2 children would eat more chocolate if one of them ate *one* part of the first chocolate, and the other ate *three* parts of the second chocolate bar. Whereas all students agreed that the two children would eat the same amount of chocolate, only Lineo was able to make adequate connections with the probability situation and recognize that indeed Item 2 (Figure 3) represented a fair game. In essence the chocolate bar episode represented a scaffolding activity that enabled Lineo to attain a conceptual breakthrough in building the more stable part-part schema needed for ordering probabilities. From Case's (1996) perspective, Lineo's thinking, at the end of the instructional period, corresponded to integrated bidimensional thinking as manifest in her ability to order probabilities in both simple and compound random situations.

4. SUMMARY AND CONCLUSIONS

This chapter has provided a detailed description of upper elementary and middle school students' ability to generate complete sets of outcomes and to make valid probability predictions for simple and compound events. Additionally, it has given an account of how upper elementary students' thinking in relation to sample space and likelihood predictions for simple and compound random experiments evolve in an instructional setting.

Looking at upper elementary students' growth in probabilistic thinking suggests the kind of cognitive background and potential that they are likely to bring to the middle school program in probability. The general picture appears to be that a carefully designed instructional sequence can enable students to finally experience consistent success at generating complete sets of outcomes for compound random processes. Students' strategies for listing sets of outcomes for simple and compound random experiments include (in increasing degree of sophistication) (a) arbitrary lists and incomplete lists based on subjective reasoning for simple events, (b) trial-and-error strategies, (c) partially-generative strategies for compound events, and (d) generative strategies for compound events.

An instructional process that appeared to foster students' proficiency in listing complete sets of outcomes for simple and compound random experiments was the creation of a problematic situation that produced a cognitive conflict and motivated students to use a more systematic process. Discussions in small group and whole-class settings constituted a supportive context for this problematic situation. It emerged that the listing of complete sets of outcomes for compound random processes entailed first using a part-part schema in order to count the number of elements in each set, before integrating these into a single structure. Furthermore, students recognized the need for the multiplication rule as a means of assessing whether their listed sample space was complete. From a psychological perspective, it appears that this complex sample space operation may only be performed by students operating at Case's (1996) integrated bidimensional level; that is by students who have constructed multiple mental counting lines and are able to perform a number of operations simultaneously and flexibly.

When upper elementary students make probability predictions in simple and compound random experiments, a number of thinking tendencies come to the fore: (a) subjective responses, (b) use of informal quantitative phrases to describe probabilities, and (c) an ability to link predictions to sample space composition and to order probabilities accordingly. As in the case of generating complete sets of outcomes, a number of factors contributed to the progress made by these students during instruction. These include (a) small-sample experimenting, (b) focusing on an analysis of sample space composition, and (c) conceptual questions posed in small group or whole-class settings. Paradoxically, most of the students experienced more success in listing sets of outcomes than in making probability predictions for compound random experiments. It appears that making predictions for compound random processes is a much more complex phenomenon: Students need to conceptually view and assess the combination of elements in the sample space before doing the counting and ordering required for making a valid prediction. Psychologically, it seems that students need to have attained a cognitive level similar to what Case (1996) termed the integrated bidimensional level, before they can make probability predictions in the case of compound events. At this cognitive level, students have constructed and are able to operate on more than one mental counting line; they are also able to perform the required arithmetic.

Another key observation is that central conceptual structures that seem to drive the development of students' probabilistic thinking are contained in part-part and part-whole schemata. In fact a number of the students in Polaki's (2002b) study struggled in their attempt to order probabilities because of their failure to make part-part and part-whole comparisons. Research on students' understanding of rational number concepts (e.g. Singer & Resnick, 1992) has indicated that these same schemata play a key role in dealing with rational numbers. The teaching of probability and rational number concepts might benefit learners to a greater extent if teachers attempt to organize instructional activities and tasks that evoke growth in students' ability to make part-part and part-whole comparisons.

Further research work in these knowledge domains is needed to explore students' growth in learning to make valid predictions for compound events.

Such a body of research has a potential to inform learning experiences and episodes that are critical to the development of students' ability to deal with various aspects of compound events.

REFERENCES

Benson, C. T., & Jones, G. A. (1999). Assessing students' thinking in modeling probability contexts. *The Mathematics Educator, 4(2)*, 1-21.

Biggs, J. B., & Collis, K. F. (1991). Multimodal learning and the quality of intelligent behavior. In H. A. Rowe (Ed.), *Intelligence: Reconceptualization and measurement*. Hillsdale, NJ: Erlbaum.

Carpenter, T. P., & Fennema, E. (1988). Research and cognitively guided instruction. In E. Fennema, & T. P. Carpenter (Eds.), *Integrating research on teaching and learning mathematics (*pp.2-17). Wisconsin: Wisconsin Center for Education Research.

Case, R. (1996). Reconceptualizing the nature of children's conceptual structures and their development in middle childhood. *Monographs of the Society for Research in Child Development, 61, (1-2, Serial No.246)*, 1-26.

Cobb, P. (2000). Conducting teaching experiments in collaboration with teachers. In A. Kelly & R. Lesh (Eds.), *Handbook of research design in mathematics and science education (pp.307-333)*, Mahwah, NJ: Erlbaum Associates.

English. L.D (1990). Young children's combinatoric strategies. *Educational Studies in Mathematics, 22*, 451-474.

English, L.D (1993). Children's strategies for solving two- and three-dimensional combinatorial problems. *Journal for Research in Mathematics Education, 22*, 255-273.

Fischbein, E., & Schnarch, D (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research in Mathematics Education*, 28, 96-105.

Green, D. R. (1983). A survey of probability concepts in 3,000 pupils aged 11-16 years. In D. R. Grey, P. Holmes, F. Barnett, & G. M. Constable (Eds.), *Proceedings of the First International Conference on Teaching Statistics* (pp.766-783). Sheffield, England: Teaching Statistics Trust.

Hogg, R. V., & Tanis, E. A. (1997). *Probability and statistical inference* (5^{th} ed.). Upper Saddle River, NJ: Prentice Hall.

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.

Jones, G. A., Thornton, C. A., Langrall, C. W., & Tarr, J. E. (1999). Understanding students' probabilistic reasoning. In L. V. Stiff, & F. R. Curcio (Eds.), *Developing mathematical reasoning in Grades K-12: 1999 Yearbook (pp.146-155).* Reston, VA: National Council of Teachers of Mathematics.

Lamon, S. J. (1999). *Teaching fractions and ratios for understanding: Essential content knowledge and instructional strategies for teachers.* Mahawh, NJ: Erlbaum Associates.

Lecoutre, M. P. (1992). Cognitive models and problem spaces in "purely random" situations. *Educational Studies in Mathematics, 23*, 557-568.

National Council of Teachers of Mathematics (2000). *Curriculum and evaluation standards for school mathematics.* Reston, VA: Author.

Piaget, J., & Inhelder, B. (1975). *The origin of the idea of chance in students* (L. Leake, Jr., P. Burrell, & H. D. Fischbein, Trans). New York: Norton (Original work published 1951)

Polaki, M. V., Lefoka, P. J., & Jones, G. A. (2000). Developing a cognitive framework for describing and predicting Basotho students' probabilistic thinking. *Boleswa Educational Research Journal, 17,* 1-21.

Polaki, M. V. (2002a). Using instruction to identify mathematical practices associated with Basotho elementary students' growth in probabilistic reasoning. *Canadian Journal of Science, Mathematics and Technology Education, 2,* 357-370.

Polaki, M. V. (2002b). Using instruction to identify key features of Basotho elementary students' growth in probabilistic thinking. *Mathematical Thinking and Learning, 4,* 285-314.

Pratt, D. (2000). Making sense of the total of two die. *Journal for Research in Mathematics Education, 31*, 602-625.

Simon, M. A. (1995). Reconstructing mathematics from a constructivist pedagogy. *Journal for Research in Mathematics Education, 26*, 146-149.

Singer, J.A., Resnick, L. B. (1992). Representations of proportional relationships: Are children part-part or part-whole reasoners? *Educational Studies in Mathematics, 23,* 231-246.

Speiser, R., & Walter, C. (1998). Two dice, two sample spaces. In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee, & W. K. Wong (Eds.), *Proceedings of the Fifth International Conference on the Teaching of Statistics (Vol. 1, pp. 1041-1047).* Voorburg, The Netherlands: International Statistical Institute.

Vidakovic, D. (1998). Children's intuition of probabilistic concepts emerging from fair play. In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee, & W. K. Wong (Eds.), *Proceedings of the Fifth International Conference on the Teaching of Statistics (Vol. 1, pp. 67-73).* Voorburg, The Netherlands: International Statistical Institute.

Watson, J. D., Collis, K. F., & Moritz, J. B. (1997). The development of chance measurement. *Mathematics Education Research Journal, 9*, 60-82.

Watson, J. D., & Moritz, J. B. (1998). Longitudinal development of chance measurement. *Mathematics Education Research Journal, 10*, 103-127.

Zimmerman, G. M. & Jones, G. A. (2002). Probability simulation: What meaning does it have for high school students. *Canadian Journal of Science, Mathematics, and Technology Education, 2*, 221-237.