

# Fairness of Dice: A Longitudinal Study of Students' Beliefs and Strategies for Making Judgments

Jane M. Watson and Jonathan B. Moritz, *University of Tasmania, Australia*

One hundred eight students in Grades 3, 5, 6, 7, and 9 were asked about their beliefs concerning fairness of dice before being presented with a few dice (at least one of which was “loaded”) and asked to determine whether each die was fair. Four levels of beliefs about fairness and four levels of strategies for determining fairness were identified. Although there were structural similarities in the levels of response, the association between beliefs and strategies was not strong. Three or four years later, we interviewed 44 of these students again using the same protocol. Changes and consistencies in levels of response were noted for beliefs and strategies. The association of beliefs and strategies was similar after three or four years. We discuss future research and educational implications in terms of assumptions that are often made about students' understanding of fairness of dice, both prior to and after experimentation.

**Key Words:** Beliefs; Children's strategies; Cognitive development; Longitudinal studies; Probability; Qualitative methods

The teaching that takes place in schools using dice to model random outcomes rests on the assumption that students believe that dice are *fair*; that is, each side has the same likelihood of “coming up.” The issue of fairness is often dismissed in classrooms with questions such as What is the chance any side will come up when this die is tossed? Hearing a chorus of “1/6” from students, teachers are likely to move on to issues they consider to be more sophisticated and more likely to challenge their students, such as what happens when two dice are tossed and the outcomes summed (e.g., Australian Education Council [AEC], 1994, pp. 134–135), or the fairness of games whose rules are determined based on the outcomes of dice tosses (e.g., Bright, Harvey, & Wheeler, 1981). Even when trials are performed for a single die, the purpose is usually to verify the fairness of the die rather than put fairness to the test, and any empirical deviations from an even distribution are likely to be dismissed as natural variation.

Today, the assumption of fairness is rarely questioned, either in classrooms or in textbooks. However, this was not always so. The ancient astragalus, which comes

---

This research was funded by Australian Research Council Grants, Nos. A79231392, A79800950, and W000910. The authors acknowledge Lionel Pereira-Mendoza for his contributions to discussions at the time of planning the protocol.

from the heel bone of a hoofed animal and is probably the earliest chance device, was definitely not fair in the sense that each of its four sides was not equally likely to occur topmost when tossed (Bennett, 1998). This was not terribly important at the time, however, because it was believed that a god or fate determined the outcome, not chance based on the shape of the astragalus. Whatever the belief in the mechanism for determining outcomes, the outcome of *winning* has been an important concept since the advent of gambling in ancient times. The desire to win led to the early “loading” of dice. Bernstein (1998) reports that “although Egypt punished compulsive gamblers by forcing them to hone stones for the pyramids, excavations show that the pharaohs were not above using loaded dice in their own games” (p. 13). By the time of Cardano and Galileo, the idea of fairness was firmly entrenched with Cardano qualifying his analysis of probabilities with “if the die be honest” (Bennett, 1998, p. 77) and Galileo describing a fair die as one with “six faces and when it is thrown it can equally well fall on any one of these” (p. 47). Today it is likely to be assumed that manufacturing quality ensures that dice are fair for all practical purposes, although Peterson (1998) notes that “ordinary store-bought dice... generally have recessed spots and distinctly rounded edges ... [being] probably somewhat biased” (p. 7) and discusses how fairness is carefully achieved in casino dice. Bernstein (1998), however, goes on to note further qualifications on fairness: “Even... throws of dice will vary in response to slight differences in the energy that puts them in motion. Unable to observe such tiny differences, we assume that the outcomes they produce are random, unpredictable” (p. 201). Using trials to test hypotheses about fairness was an idea known, at least theoretically, to Cicero and Cardano (Bennett, 1998), and Peterson (1998) relates the story of a man who, in the 20th century, confirmed the fairness of casino dice with a million trials.

The issues that students face when handling dice today are not new but similar to those faced by others throughout history. The beliefs about dice that students bring with them to school also include beliefs that God, fate, or mental powers determine dice outcomes (J. Truran, 1985; K. Truran, 1995). Such beliefs also include understandings about dice developed through experiences when playing games, such as the importance of rolling technique, experiences of losing games, and the difficulty of obtaining a 6 to start in a game (Kerslake, 1974). For many students, these intuitions about dice and probability are resistant to instruction (Fischbein & Gazit, 1984) and do not improve with age (Fischbein & Schnarch, 1997).

In the study reported in this article, we explored two aspects of student understanding of dice. The first was students’ beliefs about the fairness of dice. The second involved students’ strategies for judging the fairness of dice when presented with dice that were fair, had repeated numbers, or were unevenly weighted. Anecdotal evidence suggested to us that there may be some situations where these two aspects of understanding fairness (i.e., beliefs and strategies) may not be closely related. We identified a number of people like the mathematics major who claimed, after completing several hundred trials tossing a coin and confirming equality of heads and tails, “I know the chance of heads and tails are the same but

I always chose tails because it comes up more for me.” Beliefs may be based on intuitions, experience, or classical theoretical assumptions concerning equally likely outcomes, and may be divorced from empirical approaches of gathering data to test such hypotheses. In our study, we considered both beliefs about fairness and strategies for judging fairness, and the relationship between them. Also of interest, given the observations of Fischbein and Gazit (1984) and Fischbein and Schnarch (1997), was the stability of beliefs and strategies over time.

## CURRICULUM BACKGROUND

Questions about the fairness of random generators occur within the overall milieu of students’ understanding of the relationship between pattern and variation within random processes. As Moore (1990) states, “Phenomena having uncertain individual outcomes but a regular pattern of outcomes in many repetitions are called *random*” (p. 98). It is the intuition required to balance the uncertain and the regular that determines fairness and which Moore believes is missing for many students who are taught probability theory without experiencing random behavior in the classroom. At the most sophisticated level, questioning fairness involves being aware of classical probability theory based on equally likely outcomes as well as engaging with probability based on relative frequency in long-term empirical outcomes (Borovcnik & Bentz, 1991). Streefland (1991) claimed, however, that young children brought experience from outside of school to discussions of fractional sharing, and these led to valid intuitions of equal parts being fair. The same underlying intuition is appropriate when considering fairness of dice, and this idea appears to be behind suggestions in *A National Statement on Mathematics for Australian Schools* (AEC, 1991). It claims that upper-primary students should “make non-numerical predictions about equally likely events, such as those involved in rolling a fair die” (p. 170) and “conduct experiments with coins, dice and spinners; record and organise the data, and compare the results with predictions” (p. 170–171). During the secondary school years, students should “devise and carry out simple experiments (e.g., dice, spinners, coins, tossing a thumbtack) to estimate probabilities; compare the results with results determined by analysis, and discuss the differences observed” (p. 176). The approach here is that of comparing expected fairness, perhaps derived from theory, with empirical observations of the random generator.

The method of testing theories or expectations against trials is more commonly suggested in relation to judging the rules of games for fairness in which the fairness of the random generator is taken for granted (e.g., National Council of Teachers of Mathematics [NCTM], 1989; Ministry of Education, 1992). In the *Principles and Standards for School Mathematics* (NCTM, 2000), although the notion of fairness is not explicitly mentioned, a useful technique of collecting data to test a hypothesis is detailed for students in Grades 6–8:

[Students] could discuss whether the results of the experiment are consistent with their predictions. If students are accustomed to reasoning from and about data, they will

understand that discrepancies between predictions and outcomes from a large and representative sample must be taken seriously. The detection of discrepancies can lead to learning when students turn to classmates and their teacher for alternative ways to think about the possible results. (pp. 254–255)

This approach, whether applied to the use of fair random generators given certain rules in a game or to the fairness of the random generator themselves, is one of the general goals of the school mathematics curriculum today.

## RESEARCH BACKGROUND

Over the years, various researchers have considered students' beliefs about fairness of dice (e.g., Amir & Williams, 1999; Green, 1983; Kerslake, 1974; Lidster, Pereira-Mendoza, Watson, & Collis, 1995; Lidster, Watson, Collis, & Pereira-Mendoza, 1996; J. Truran, 1985; K. Truran, 1995; Watson, Collis, & Moritz, 1997), thus indicating that many students have difficulty with the idea. In the early primary grades, Kerslake (1974) found that approximately 30% of students in those grades believe dice are fair, whereas by secondary school, Green (1983) found that between 67% and 86% of students hold this belief. Some students believe the apparently contradictory ideas that all numbers on a die have the same chance of occurring, but that the number 6 is least likely to occur (Konold, Pollatsek, Well, Lohmeier, & Lipson, 1993). Kerslake (1974) commented that children's experiences playing games that have 6 as a special outcome desirable for some purpose may contribute to the belief that 6 is less likely to occur. She suggested research on the extent to which "the experimental approach of throwing a great many dice ... convinces children that the probability of each score is  $1/6$ " (p. 22).

Using simulated results presented in graphs, Green (1983) asked secondary students which of three graphs might be the outcome of 60 throws of a fair die: A graph showing frequencies between 8 and 12 for each number was chosen by 47% of the students, a graph showing higher frequencies for the middle numbers by 36% of students, and a graph showing frequencies of exactly 10 for each number by 17% of students. Watson and Chick (2001) used a similar technique to examine students' beliefs about fairness of dice in a collaborative setting. Groups of three students in Grades 3, 6, and 9 played games with dice, some of which were loaded, and then were shown a series of labeled graphs that simulated frequencies for samples of 60 throws, 360 throws, 600 throws, and 12,000 throws for five dice, some of which were unfair. Students were asked to decide whether the sample results shown in each graph were likely to represent outcomes of fair or unfair dice. Some students seemed to conceive of fairness as applying to a single number by comparison of frequencies of that number across different dice, rather than to a single die by comparison across different numbers for that die. As one student said, "1 is probably the fairest number." Some students tenaciously clung to the idea of fairness, even when shown graphs that represented an uneven distribution. In the words of another student, "I reckon all the dices are fair, every single one of them [...] every single dice that I've thrown [...] it's the way you throw the dice, that's the reason

why I think they are all fair.” Other students understood that dice are normally fair, but qualified this with other comments related to (a) how the dice are rolled, (b) the appropriate weight distribution, or (c) possible variation of sample results from the uniform distribution expected for normal, fair dice. These three ideas are illustrated by the following student comments: (a) “If you rolled it that many times it would come out all exactly the same [even distribution], *unless you rolled it exactly a special way* [emphasis added] so it would come out exactly the same every time”; (b) “If it happened three times in a row and it’s getting further and further ahead, then *it’s weighted or something* [emphasis added]”; (c) “The dice would have to be perfect to get that [exactly even distribution] ... because *normally you don’t get exactly the same* [emphasis added].” Lidster et al. (1996) also noted that some students apply the concept fairness to several numbers on a die in comparison to others or believe that dice may be considered fair in relation to each other if they behave in a similar fashion.

Although the balance of expectation of pattern (e.g., equally likely outcomes) with variation (e.g., short-term runs) has been studied for random processes for coin tossing and the distribution of points in a two-dimensional array (Batanero & Serrano, 1999; Green, 1986, 1991), students were asked to make judgments about outcomes presented to them, but not asked to collect data for themselves as is recommended in curriculum documents (e.g., AEC, 1991). In the study reported here, we were interested in how students judge the fairness of a random generator (e.g., a die), without the stimulus of a series of observed outcomes but with the potential to generate their own outcomes. Lidster et al. (1995, 1996) reported on some early observations of a subset of the data as well as on the data from students collaborating on comparing graphs of simulated outcomes (see also Watson & Chick, 2001). They found that students in Grades 3, 5, 6, and 7 suggested strategies for determining fairness that may be described broadly as either *observational* or *empirical*.

Observational strategies included noting physical characteristics of a die such as the presence of one of each number on the faces or whether the shape of the die was a perfect cube and was evenly weighted (e.g., “the missing bits where the dots are might make it unfair,” [Lidster et al. 1995, p. 10]), discussing the roll being random (e.g., “you can make the dice unfair by the way you roll it,” [p. 11]), and combinations of these characteristics (e.g., “if all the numbers are there and the edges are even then the dice is fair,” [p. 11]). Similar strategies were found by Ritson (2000), who repeatedly interviewed upper primary grade students, and in some interviews asked students whether a normal die and a cuboid measuring 10 mm × 15 mm × 15 mm would both yield the same chance for all numbers. Two students were quoted as claiming that both dice would be fair because all numbers (i.e., possible outcomes) were present, even after noticing the difference in the shape of the objects. A third student, however, noted, “It wouldn’t be the same chance because some of its sides are different” (Ritson, 2000, p. 14).

Empirical strategies for judging fairness depend on a number of understandings and skills of data handling. Students must appropriately design and record the data

collection, such as recording which features are relevant to record (AEC, 1991). They must also understand the tension between representativeness and variability (Metz, 1997), in particular that increasing sample size increases the confidence that empirical frequencies reflect the theoretical probabilities of the random generator. Many students believe, however, that outcomes in the short-run should represent the equally likely outcomes of the random generator (Shaughnessy, 1992). Such beliefs affect the interpretation of results and discernment of what variation would be expected or surprising enough to question a prior hypothesis. Lidster et al. (1996) observed different levels of sophistication in responses of students who appreciated the need to do trials to determine fairness. For example, suggestions were made by primary students who described playing games all day and by secondary students who described conducting 1,000 trials and keeping a tally. When presented with graphs showing outcomes of trials for fair and unfair dice, some primary students compared samples across trials but failed to appreciate the significance of a larger sample size, whereas many of the ninth-grade students realized the importance of sample size and based their decisions on the graphs of the largest sample size.

Students' longitudinal development of probabilistic concepts was first considered by Green (1991). He surveyed 305 students at ages 7–10 years and again 4 years later, on questions related to *randomness* and *comparison of odds*. In terms of randomness, there was virtually no change over the 4 years in prediction of random outcomes but for comparison of odds, improvement was quite dramatic. Green suggested the difference for the two topics may have been related to ratio being a significant topic in the mathematics curriculum over these years, whereas random behavior of generators was not. Other longitudinal survey work on students' understanding of chance measurement was carried out by Watson and Moritz (1998). Over a 4-year period, they observed significant improvement in performance for about 200 students on items about belief in the equal likelihood of dice outcomes, about drawing names from a hat, and about comparing odds in items similar to those in Green's study.

These longitudinal survey studies were useful in monitoring change but did not explore student understanding in nontest conditions or where concrete materials could be used to allow experimentation. Green (1993) echoed the call of Garfield and Ahlgren (1988) for "longitudinal studies of how individuals actually develop in stochastic sophistication" (p. 58). A start in the documentation of development over time for individuals has taken place within the larger project of which the current study is a part. Watson and Moritz considered the longitudinal development of ideas associated with average (2000b) and with representing, interpreting, and predicting with pictographs (2001a). Watson (2001) also followed the longitudinal development of inferential reasoning and observation of variation in graphical presentations. The study reported here is another step in understanding how students' concepts change over time.

Our investigation reported here involved two studies to document the development of students' beliefs and strategies in dealing with the fairness of dice. Study 1

addressed the following questions in relation to tasks about fairness of dice for students in a cross-section of grades.

1. What do students believe about the fairness of dice? What experiences or understandings do students state to support their beliefs? Do these show a hierarchical progression? Do beliefs differ for students of different grades?
2. What are students' dominant strategies for assessing the fairness of dice? What are the qualitative differences among these strategies, and do they fit within a framework of increasing sophistication and statistical appropriateness? Do strategies differ for students of different grades?
3. Is there an association between students' beliefs in the fairness of dice and their strategies for assessing the fairness of dice

Study 2 involved observing longitudinal change in some students who participated in the first study, and addressed this question:

4. How do students' responses change over time (i.e., 3 or 4 years) to questions about belief in the fairness of dice and about assessing the fairness of dice? Is there a change in the association between beliefs and strategies?

## METHOD

### *Participants*

In Study 1, interviews were conducted with 108 students in Grades 3, 5, 6, 7, and 9 from two locations, Tasmania and South Australia. Information about the number of participants in each grade level (some of whom also were in the sample for the second study) appears in Table 1. Tasmanian coeducational government schools were selected in urban and rural areas to represent Tasmanian students, and a private girls' school in South Australia was a convenience sample due to interest of the school in participating in educational research. Grade levels were selected to cover a cross-section of grades and to include the highest grade level in primary schools (i.e., Grade 6 in Tasmania and Grade 7 in South Australia). Because the

Table 1  
*Information About the Participants in Study 1 and Study 2*

Grade	Tasmania		South Australia	
	Study 1	Study 2	Study 1	Study 2
3	26	8	6	5
5, 6, 7 <sup>a</sup>	26	10	16	12
9	26	4	8	5
Totals	78	22	30	22

<sup>a</sup> Because of the small number of students in each grade-level sample and other factors (detailed in the article), we combined the grade levels to form a middle-school sample of students.

*Note.* Tasmanian students in Study 2 were interviewed after 4 years and their counterparts from South Australia were interviewed after 3 years.

purpose of the study was to describe development of understanding across grades and across time, the total sample provided data from all sectors of schooling in Australia. These students were part of a larger project about stochastic understanding, and they were selected for interviews on the basis of the variety and sometimes the unusual nature of their responses to survey items about chance and data (Watson, 1994). Selections were not necessarily based specifically on responses related to dice; other tasks administered in interviews were based on comparing data sets, sampling, average, and pictographs (Watson & Moritz, 1999, 2000a, 2000b, 2001a). Teachers confirmed that the students selected by the researchers were articulate and willing to be interviewed.

To address Research Question 4, Study 2 involved follow-up interviews of 44 of the 108 students from Study 1. The numbers of those students reinterviewed are given in Table 1. In Tasmania, we interviewed the students 4 years later and in South Australia, 3 years later. The differing gap in time was due to funding constraints of the subsequent research that allowed Study 2 to take place. Analysis of other data (Watson & Moritz, 2000b) suggested that the long time intervals and lack of evidence of specific instruction related to the topic being studied would lead to similar conclusions about developmental change. The retention rate was higher in South Australia because most of the students were in the same private school 3 years later and all 22 students agreed to be interviewed again. In Tasmania, all students from Study 1 had changed schools in the intervening years and for those in Grade 9 in Study 1 it was only possible to trace those who had subsequently enrolled at the state's university.

Students in both Tasmania and South Australia experienced mathematics curricula influenced by national curriculum documents (AEC, 1991, 1994) but there was no mandated curriculum in either state and no intervention by the researchers during the interval between interviews. Anecdotal evidence suggested that whereas the socioeconomic status of the South Australian students was likely to be higher than that of Tasmanian students, the classroom instruction was likely to be similar in each state. These smaller groups in Study 2 each appeared, from the distribution of responses to other tasks, to be representative of the original group of students, except for the four Grade 13 students from the Tasmanian sample, who were the Grade 9 students in Study 1 (see Watson, 2001; Watson & Moritz, 2000b). These four students had chosen to continue their education at the university level, and hence were likely to be of higher overall ability than the rest of their cohort in Study 1.

### *Materials and Interview Protocol*

Students were given a number of wooden dice with 3 cm edges. The dice were typical in that they had "dots" on each face representing numbers (e.g., one dot representing the number 1). For Study 1, students had three dice: a Red one that was theoretically fair and with the numbers 1 through 6 represented; a White one with only the numbers 1, 2, and 3 each represented twice (opposite faces had the



same number of dots); and a Blue one, again with the numbers 1 to 6 represented, which had been weighted on the side with the 5 dots. The Blue die was configured with opposite faces summing to 7, and the trials ( $n = 200$ ) we conducted yielded the following distribution of outcomes: 1 (16%), 2 (33%), 3 (20%), 4 (16%), 5 (7%), and 6 (9%). In first six interviews for each of Grades 3 and 6 in Study 1, the White die was not yet available. Because the distribution of levels of response for these students was similar to those who had all three dice, data from all students were included in the study. For Study 2, there was also a Green die available, weighted in a similar way to the Blue one.

The interview protocol used in this investigation began with the placing of the dice on the table in front of the student and asking an introductory question, such as “Do you play games with dice?” Next, a question was asked to elicit students’ beliefs about fairness of dice. Sometimes a frequency form of the question was used, similar to “Do some numbers come up more often than others?” and sometimes a chance form was used, such as “Do all numbers have the same chance of coming up?” The interviewer then asked for clarification, either of unfair beliefs by asking which numbers occur more often, or of fair beliefs by asking for confirmation that all numbers come up equally often *and* that they have the same chance.

As a transition into considering strategies for determining fairness, the question “Do you know what it means for dice to be fair?” was used. This was for clarification only and if the students were unclear, we made a statement like the following: “We say a die is fair if all six numbers have the same chance of coming up.” The interviewer then asked students how they would work out which of the dice on the table were fair and which were unfair. Students who failed to inspect or throw the dice were told that they were permitted to do so. Students who continued to assert beliefs without engaging the dice in front of them were further asked what they might do to be *sure* or to demonstrate they were right to be skeptical. In interviews where students used observation alone to conclude that the White die was unfair and the other dice fair, students were sometimes told some of the remaining dice in front of them were fair and some were unfair, and were asked to judge their fairness. The wide variation in initial responses necessitated flexibility on the part of the interviewer, and the excerpts that follow include interviewer prompts to clarify the part these may have played in the students’ responses.

### *Procedure*

Students were interviewed individually in a separate room for 45 minutes during class time by one of the authors; all interviews were videotaped. Students were told that they could stop the interview at any time, but none chose to end the interview early. The questions described above constituted a protocol that was one of nine protocols used during the interview, covering various aspects of the chance and data curriculum. The fairness-of-dice task appeared early in the interview, thus students had little expectation of how long to spend on the task; some students believed the fairness could be judged within a minute, whereas others devoted over 10 minutes

to recording results and interpreting them. The interviewer encouraged the student to spend as much time as needed on the task. If the student suggested more than 30 rolls of each die, the interviewer agreed that a large number was a good idea, but suggested the student do fewer rolls to suit the time constraints of the interview. At the end of the protocol, students who found the dice to be unfair were shown the loaded die or dice, and the process of creating the bias was described; they were assured that such bias was not normal.

### *Analysis*

Digital video segments and transcripts of the dialogue were hypertext-linked from a spreadsheet that formed a tool for analysis of responses, thus allowing us to have easy repeated access to the raw interview data. We both read transcripts and viewed the video data of all student responses independently using the spreadsheet to record observations. For the first stage of analysis, preliminary column headings of response features were suggested by an earlier reading of the South Australian transcripts. As we examined the entire response set, we identified additional features of responses occurring across a number of responses, which necessitated the introduction of additional spreadsheet columns to record these features. After consultation, we agreed on the coding of features of responses recorded in these spreadsheet columns. In particular, our discussion involved the point in the dialogue where students appeared to conclude their responses of their own initiative. Further dialogue was then dependent on the interviewer prompting the students for further engagement with the task. Codings were decided based on students' initial responses but at times further dialog is reported here because of potential interest in the reluctance of students to engage in experimental trialing. In the second stage of analysis, responses that exhibited common patterns of these features were then clustered into categories (see Miles & Huberman, 1994), based on a combination of the statistical appropriateness of the comments and their structural complexity. These two stages were performed in a few cyclic iterations (Miles & Huberman, 1994, p. 61) to improve reliability of coding.

The identification of different beliefs and strategies was informed by a combination of (a) the statistical appropriateness of the students' comments and (b) their structural complexity. Statistical appropriateness included appreciation that dice should have equally likely outcomes, subject to a degree of random variation when trialed, and knowledge of strategies for examining and empirically trialing dice outcomes (AEC, 1991). Structural complexity was informed by the cognitive development model of Biggs and Collis (1982, 1991), which had been used with other response data from students also included in this study (e.g., Watson & Moritz, 1999, 2000a, 2000b, 2001a). This model incorporates various modes of thinking similar to Piagetian stages, including the *ikonik* mode, involving intuitions or storytelling about experiences, and the *concrete symbolic* mode, involving symbols and propositions referring to concrete objects, as often taught in schools. Within each mode, hierarchical learning cycles are based on the structure of the

observed learning outcomes (or SOLO) taxonomy (Biggs & Collis, 1982) using three levels:

1. *Unistructural (U)* responses that employ single elements, sometimes creating contradictions that are not acknowledged.
2. *Multistructural (M)* responses that employ multiple elements, usually in sequence, sometimes recognizing but not resolving conflicts among the elements.
3. *Relational (R)* responses that relate elements and create closure for the task.

The previous work of Lidster et al. (1995, 1996), who adapted the general developmental model to propose a hierarchical structure for children's beliefs about fairness of dice, informed our initial discussion and decision-making on the hierarchical nature of responses. For the purposes of our study, we identified four levels of *beliefs* within students' responses—*Ikonic*, *Unistructural*, *Multistructural*, and *Relational*—which reflected increasingly more appropriate statistical understanding. Students' beliefs about the fairness of dice were assigned one of the following levels with descriptors: *Ikonic-Unfair* (IK-Unfair) in which students expressed the belief that certain numbers are favored, often basing their belief on their own experiences with dice when playing games; *Unistructural-Fair* (U-Fair) in which the belief was simply stated (or agreed to) as a proposition that all numbers have the same chance; *Multistructural-Fair qualified* (M-Fair qualified) in which students made additional comments about conditions to ensure fairness; and *Relational-Short-term variation* (R-Short term) involving contrasting comments that short-term outcomes may appear unfair but in the long term chances are fair. These levels are further summarized and explained in Table 2. In this study,

Table 2  
*Levels of Belief About Fairness of Dice*

Level	Descriptor	Explanation
Ikonic	Unfair (IK-Unfair)	Beliefs that dice are unfair, often involving stories about games. <i>Idiosyncratic</i> —belief that specific numbers are more likely to occur. <i>Inconsistent</i> —belief that some specific numbers occur more often than others <i>and</i> simple agreement that all numbers have the same chance.
Unistructural	Fair (U-Fair)	Theoretical belief in fairness or equal chance in a propositional form; any reference to experience is secondary to the main conclusion of fairness.
Multistructural	Fair qualified (M-Fair qualified)	Belief that dice are fair subject to the <i>rolling condition</i> of an unbiased rolling technique or to the <i>physical condition</i> of the manufacture of the dice for even shape and weighting.
Relational	Short-term variation (R-Short term)	Belief that outcomes are fair in the long-term, but that short-term outcomes or selective recall of experience may suggest otherwise.

we observed contradictions within responses at the IK-Unfair level where idiosyncratic views were predominant but sometimes students had no qualms about stating in a repetitive fashion, “all dice are fair.”

Students’ strategies for judging the fairness of dice were assigned the following levels: *Ikonic-Idiosyncratic* (IK-Idiosyncratic), in which inappropriate idiosyncratic strategies were suggested; *Unistructural-Untestable* (U-Untestable), in which the single belief of fairness made any other consideration of dice characteristics unnecessary; *Multistructural-Observational* (M-Observational), in which physical features of the dice were confirmed, sometimes from a few rolls; and *Relational-Empirical* (R-Empirical), in which dice were rolled to test whether the distribution of outcomes was even for each die. These levels are further detailed in Table 3. It is important to stress that the classification of responses in levels was based on our analysis of observed responses and not on what we might have thought the students intended to say but did not.

Table 3  
*Levels of Strategy Use for Determining Fairness of Dice*

Level	Descriptor	Explanation
Ikonic	Idiosyncratic (IK-Idiosyncratic)	Incorporation of intuitive beliefs (e.g., anthropomorphism) often about specific numbers (luck) and thus engagement with dice in an idiosyncratic way.
Uni- structural	Untestable (U-Untestable)	Assertion that dice are fair; no test is necessary.
Multi- structural	Observational (M-Observational)	Observation of physical features, e.g., whether each <i>possible outcome</i> is represented on one and only one face; <i>symmetry</i> in terms of measurement of weight or shape, or specified rolling technique; or a few <i>unsystematic trials</i> to support an argument but with no systematic recording technique, and results not used to draw conclusions.
Relational	Empirical (R-Empirical)	Systematic trials of the dice, recording of results, and comparison of frequencies of outcomes to consider if the distribution of outcomes is even. <i>Small sample</i> trials include fewer than 18 rolls of each die, record of outcomes, sometimes as a sequential string of outcomes. <i>Large sample</i> trials include more than 18 rolls of each die, record of outcomes, and summaries of frequencies of each outcome.

In the following presentation of results, we provide summary tables for information on the changes in levels of response across grade levels and longitudinally. The middle school grades (Grades 5, 6, and 7) were combined for these summaries due to the small numbers in those levels, the elementary school setting for all three grades, and the gap of a year between this group and each of the other two (i.e., Grades 3 and 9). Using the scores 0 through 3 for levels of beliefs and strategies, analyses of variance were carried out across the three grade level groupings to give

an indication of differences, and the Pearson product moment correlation coefficient was calculated for the beliefs and strategies variables.

## RESULTS

We present the results in the order of the research questions. Responses selected from Study 1 to illustrate response levels for Research Questions 1 and 2 were commonly chosen from those offered by students who later participated in Study 2, in order further to illustrate change in responses for Research Question 4. Unique identification of students (e.g., S1) is used throughout and some identification labels are repeated because these students have multiple responses quoted throughout this section. The use of “...” denotes a pause in the student’s response, and “[...]” denotes dialogue that has been edited, which does not affect the meaning of the extract. “I” denotes the interviewer comments, which have in some cases been edited for brevity and appear in italics within the student comments.

### Research Question 1: Beliefs About Fairness of Dice

Table 4 shows a summary of the levels of response for the 108 students in Study 1 by grade level for beliefs about fairness of dice. Ninety-four students responded with beliefs that were classified as either IK-Unfair or U-Fair. Only 14 students gave responses that qualified fairness (M-Fair qualified) or commented about short-term variation (R-Short term). Students in Grade 9 were less likely than those in lower grades to respond at the IK-Unfair level. Although it is possible to claim that the mean level of response was higher as grade level increased ( $F(2,105) = 4.18, p = .018$ ), adjacent grade groupings were not significantly different. In the next sections, we present examples of responses to illustrate the descriptions given for the four levels.

#### *Ikonic-Unfair Beliefs*

Forty-four students expressed beliefs that dice are unfair. We identified two response categories—idiosyncratic and inconsistent—depending on whether students offered more than one suggestion.

Table 4  
*Frequency of Level of Belief by Grade*

Level of belief	Grade			Totals
	3	5–7	9	
IK-Unfair	20	17	7	44
U-Fair	9	20	21	50
M-Fair qualified	3	2	5	10
R-Short term	0	3	1	4
Totals	32	42	34	108

*Idiosyncratic.* Thirty-eight students stated beliefs that some specific numbers occurred more often than others; that is, they had more “chance.” Some responses, like the ones that follow, were apparently based on remembered personal experience or other ideas that were idiosyncratic for particular students:

S1: (Grade 6) Yeah, 3s, 4s, and 2s.... Or the ones that you don’t want to come up.

S2: (Grade 3) Yes. 6 and 5 for me.

Other responses were based on specific experiences (e.g., games) from rolling dice with a focus on getting the number 6. Focusing on this outcome appears to have distorted the students’ perceptions of the outcomes that occurred for *all* rolls:

S3: (Grade 9) 6s don’t come up as often as smaller numbers. [*I: So that’s when you are playing games?*] Yes, especially when you want a 6 to start.

S4: (Grade 5). Yes. My mum always gets 6s and I always get 2s or 1s. [...] My hand just must be not very good at throwing the dice.

S5: (Grade 9) Yes. [*I: Which ones come up more often?*] 1.

Finally, in one response a student appeared to show confusion concerning “which number is most likely to occur” and “which number is the highest number”:

S6: (Grade 3) 6. [*I: Do you think a 6 comes up more often?*] That’s the most in the dice. Well, it’s the most number in the dice.

*Inconsistent.* Six students responded with some evidence of fairness, but also with other comments suggesting certain numbers were favored. The following response reflected both fair and unfair views, as well as a conflict between fairy-tale beliefs and personal experience:

S7: (Grade 3) Yes, number 3 because in fairy-tales there’s 3 wishes and there’s 3 fairy god-mothers and there’s 3 wishes and there’s all sorts of 3 things. [*I: ... do you find that 3 comes up a lot, or do all of the numbers come up?*] Well 6 comes up not the most because it’s the biggest, and for me 2 comes up usually. [*Interviewer: Do you think that they all have the same chance of coming up?*] Yes, I think that’s just because you turn the dice differently every time.

The following ninth-grade student believed the numbers *occur* equally frequently, however, a follow-up question by the interviewer concerning the *chances* of the numbers occurring revealed inconsistent beliefs. In contrast to the third-grade student (S7) above, this student appears to acknowledge the ambivalence:

S8: (Grade 9) No, not really. [*I: So would you say that they all have the same chance, would you?*] [12 second pause] Well, 6 and 1 seem to come up less, so no not really.

### *Unistructural-Fair Beliefs*

Fifty students simply stated, or agreed to the statement, that no numbers were more likely to come up than others and hence all numbers have the same chance on fair dice. The basis of this understanding was often not stated, that is whether it was based on personal experience or on a theoretical belief. These responses are typical of the concrete symbolic mode in which propositions are stated referring to concrete situations and not restricted to telling of isolated stories as at the ikonic level:

- S9: (Grade 3) [5 second pause] [*I: Or do you think they come up about the same?*] They all come up the same.
- S10: (Grade 9) No, I think that they all have the same chance of getting rolled than all of the rest of the numbers.

### *Multistructural-Fair Qualified Beliefs*

Ten responses were more complex structurally as students volunteered qualifications about the fairness of dice. Responses included requirements for fair dice outcomes, such as how the dice are rolled or physical characteristics of the dice.

*Rolling condition.* Four students stated that no numbers were more likely to come up than others, but that this depended on an appropriate rolling technique. One student, for example, commented that minimal thrust when rolling might introduce bias in the outcomes, and that it is possible to exploit some aspects of rolling technique to bias outcomes:

- S11: (Grade 5) Not really. [*I: They all have about the same chance, do you think?*] Yes. It depends how you roll them. [*I: Right, how does it depend on that?*] [S manipulates Red die.] If you roll it so that you roll it that way, it could land on any number along there, but if you rolled it along that way, it could land on any of those numbers.

*Physical condition.* Six students stated that no numbers were more likely to come up than others, provided the dice were manufactured appropriately, but they also suggested that imperfections in shape or weight distribution would affect outcomes. A few students were very sensitive to these conditions, believing that dice manufactured with grooves for dots to denote the number on the face would result in an uneven weight distribution:

- S12: (Grade 7) Well, if they had huge holes [manipulates Blue die while talking], then 6 would probably come up the most because 1 is set heaviest and it's got the least numbers, but this looks like it won't happen that much ... and that they're both random.

### *Relational-Short-Term Variation Beliefs*

Four students suggested that dice are fair but that it may appear that some numbers come up more often when dice are tossed. These contrasting ideas were resolved in favor of belief in fairness, but were related to acknowledgment of apparent unfairness. This unfairness was attributed to short-term observation or selective recall. These students (e.g., S13 and S14 below) dealt with a conflict of sometimes-observed short-term results and the understanding of long-term trends, and related the ideas in an appropriate fashion.

- S13: (Grade 7) No. Umm ... it sometimes seems that the other person always gets a 6, but it's just the luck of the draw really. It's just that you might get a row of 6s, but then probably in the next game you'll get a row of 1s [laughs]. And, but usually it just depends, sometimes it might come up more often, but then the next time it won't. It's just the luck of the draw.
- S14: (Grade 6) Well, if it was, if you had to say, sort of like the chance, you would say

that all numbers have the same chance, but sometimes it doesn't turn out that way. Because sometimes we do, in our class, we have things that we, we make our own dice, and then we have to roll them 60 times, and see which comes up the most, out of all the numbers.

### Research Question 2: Strategies for Assessing Fairness of Dice

When asked to judge fairness of specific dice, students engaged with the dice in four ways: with idiosyncratic strategies, with “untestable” strategies (based on the belief that dice are fair), with observational strategies, and with empirical strategies involving trialing the dice. As shown in Table 5, most third-grade students responded with either idiosyncratic strategies (IK-Idiosyncratic) or observational strategies (M-Observational), whereas for students in Grades 5–7 and Grade 9, the modal response level involved observational strategies. Very few students in any grade level used empirical strategies (R-Empirical). There was some tendency for students in higher grades to respond at higher levels ( $F(2,105) = 12.78, p < .001$ ), with significant improvement in mean level from Grade 3 to Grades 5–7 ( $t(72) = 3.43, p < .001$ ), but not from Grades 5–7 to Grade 9 ( $t(74) = 1.76, ns$ ).

Table 5  
*Frequency of Level of Strategy by Grade*

Level of strategy	Grade			Totals
	3	5–7	9	
IK-Idiosyncratic	15	7	2	24
U-Untestable	5	4	5	14
M-Observational	12	30	21	63
R-Empirical	0	1	6	7
Totals	32	42	34	108

### *Ikonic-Idiosyncratic Strategies*

Twenty-four students attached special significance to individual number outcomes or to combinations of outcomes across two or more dice. In a few cases, the presentation of the dice with certain numbers facing up was the basis on which students judged fairness, as was the case for the following student:

S15: (Grade 5) [White die shows 2, others show higher numbers.] This one [White] is a bit unfair because you don't get much like these two [Red and Blue]. [*Interviewer turns dice so all show 2.*] [...] [*I: Does that make any difference?*] Yes. That is fair if they were all on 2.

This student did not perform any trials, as was the case for the following student who used the numbers facing up on each die compared to her beliefs about the numbers:



S16: (Grade 6) [Dice showing Red 6, White 2, and Blue 5] Well, I reckon number 6 isn't fair, because it just doesn't really come up as much as the others. I just think that it doesn't come up. I reckon a 2 does. I reckon it sort of lands on it most of the times. And the 5, well, it's in the middle really.

When prompted, this student acknowledged the White die would be unfair due to the repeated numbers, but when further prompted, did not consider trialing the others:

S16: [*I: Suppose that someone came along and they said, "I think that one of these two is unfair, it seems to keep coming up on one of the numbers a lot." Could you test out if they were right?*] No, not really, you couldn't really say that there wasn't. It's probably just a lucky chance that it came up on that one all the time.

### *Unistructural–Untestable Strategies*

For 14 students, beliefs strongly influenced their responses, and they relied solely on these beliefs for judging fairness. Of the following two students, the first had earlier stated a belief that dice would be fair, whereas the second had stated an ikonic belief:

S10: (Grade 9) [Holds all dice, casually turning, 10 second pause] I don't know. [*I: Suppose someone came along and said, "I think one of these might be pretty unfair, it just seems to keep coming up on just a few of the numbers."*] [Picks up all dice, 7 second pause] They all look pretty fine to me.

S5: (Grade 9) [*I: ... which ones might be fair, coming out evenly, and which ones might be unfair?*] [Doesn't pick up dice] They would all be the same.

When prompted about checking over the faces of the dice, these students could identify the repeated numbers on the White die, but used only this strategy provided by the interviewer for judging fairness of the other dice initiating no strategies of their own.

### *Multistructural–Observational Strategies*

Sixty-three students focused on physical characteristics of the dice in front of them, either by observation or by manipulation. The three categories identified at this level involved (a) possible outcomes—that is, observing that each of the six numbers appeared on one and only one face, (b) feeling if the dice were weighted symmetrically or not, or (c) rolling the dice to observe rolling irregularity rather than the outcome of the roll. None suggested on their own initiative to trial the dice (albeit it in an unsystematic manner) and record results to draw conclusions about fairness, although some did so after prompting.

*Possible outcomes.* Twenty-one students judged fairness according to whether each of the six numbers appeared on one and only one face. Students thus judged the White die with repeated numbers as unfair, whereas the other dice were judged as fair without any other considerations. Following closure on this decision, further questioning prompted 12 students to perform some trials, whereas 9 students continued to be uninterested in doing so:

S17: (Grade 9) This [White] isn't a fair die. [...] Because it doesn't have 4, 5, or 6 like these two [Blue and Red]. [...] Those two are fair and that one isn't.

A few students suggested that the White die with two faces each of 1, 2, 3 was still fair because it had two of each number.

*Symmetry.* Twenty students focused on physical symmetric characteristics of the dice instead of the numbers on the faces, such as the importance of the cubic shape being exact or the weight distribution being even. After judging that closure had occurred with regard to the fairness issue, the interviewer prompted the student further, asking for other ways to demonstrate the conclusion to people who might disagree. Following such a prompt, only four students trialed the dice in any systematic way, one student only after repeated prompting. Ten students could not be prompted to trial the dice, with only a few rolling the dice to demonstrate the uneven weight by the bias in the roll. The following dialogue between the interviewer and a sixth-grade student illustrates the belief that weight is the only possible strategy for judging fairness:

S18: (Grade 6) [Manipulates each die] Those two [White and Red] are fair, and that one [Blue] is unfair. [...] By putting them on an angle and seeing if they swung around. You could feel the weight, that it was at the top, like that [demonstrates swinging]. With the other ones, they didn't swing or anything. [...] [*I: Is there any other way that you could do it other than letting them feel it?*] Not really, because it's just chance. With any dice, it's chance if it lands on one number anyway, so that could happen with a normal dice as it would with a weighted dice. [*I: Do you think it's more likely to come up on some numbers than others?*] Yes. [*I: Would you be able to work which ones? Is there a way of doing that?*] Yes, by the opposite side of the weight.

Six students combined ideas of possible outcomes and symmetry to cover both types of physical characteristics. A ninth-grade student, for example, first noted the difference in weight, but then in manipulating the dice stumbled across the repeated numbers. The fact that many students identified one or other physical aspect and that few students considered both probably reflects in part the motivation of most students, who having identified one strategy, felt that this was sufficient for the task.

*Unsystematic trials.* Twenty-two students rolled the dice to demonstrate that a variety of outcomes occurred, but did not employ any systematic testing and recording to conclude that these outcomes were equally frequent. Eight students rolled the dice together, as if the task were to decide whether all three dice would together yield a fair distribution of outcomes, rather than whether fairness applied to each individual die:

S19: (Grade 3) [Rolls each die once, then once again, and considers all dice together] 6 came up more times than all the others. [*I: Did it? How many times did you throw them?*] I threw them twice, and 6 came up twice, and the other numbers only came up once. [*I: Right. Could that help you decide whether they were fair or not?*] They're unfair.

Fourteen students rolled the dice and considered each independently, sometimes with a few rolls, but had no system for collecting data, recording them, or summa-

rizing the results. Some of these students appeared to be observing if a variety of outcomes occurred rather than documenting the frequency distribution. The following student, for example, first commented about the results across the three dice, but then appeared to draw conclusions across the three trials of each die, based on which had repeated outcomes:

S20: (Grade 6) If you roll them each 3 times. [Rolls Red die once, then Blue and White] Well, I suppose that's a kind of... they're all different, so they've all got a chance of.... [Rolls each once more] They're all still different again, except 1 keeps coming up [points to White die]. [Rolls each once more] That one [Red] is the fairest because all different ones come up, and these two have had the same [Blue and White].

Other students were also unsystematic in trialing, simply rolling and observing outcomes without planning how many rolls to do or recording the results. One sixth-grade student, for example, also drew a conclusion from a few trials based on repeated outcomes; she concluded the Blue “seems to come up on the same numbers most of the time,” the White and Red dice “usually come up on different ones,” and hence White and Red “are more fair.” Other students using unsystematic trialing recorded results, but appeared to have no way of using them to draw a conclusion about the fairness.

### *Relational-Empirical Strategies*

Seven students suggested trialing the dice, and they did so with appropriate recording methods to use empirical data concerning the frequency distribution of outcomes for each die to judge fairness. It is interesting to note that of the seven students who responded in this level, none checked the numbers on the dice before trialing, and only two checked them after their trials, whereas the other five required prompting by the interviewer to check the dice and notice the repeated digits on the White die. The physical device hence mattered less than the outcomes, which were the basis for the decision. Of the seven students, four used only a small number of trials (e.g., fewer than 18 rolls of each die) and the rest used more trials.

*Small number of trials.* Four students who performed trials only used 6 to 12 rolls of each die. This may have been a function of the perceived time available in the interview because some students asked how much time they could have. Some had simple recording strategies based on memory or a sequential list of outcomes. The following student, for example, decided to do 12 trials, wrote the numbers 1 to 12 in a column, wrote the outcomes for each roll of the Blue die in an adjacent untitled column, and wrote the outcomes for the Red and White dice in columns titled by color. She also hinted at an idea about short-term variation in anticipating results, though her belief quoted earlier concerned physical conditions for fairness:

S12: (Grade 7) Well you could roll it say 12 times and then expect around each number to come up twice, but then if... it's okay to be not, to be unfair, because it's just chance like, maybe a 1 might not turn up all the time, that time. But then maybe it would turn up a lot another time. [...] Right, so that one was a 6 [records Blue results] 3, 2, 3, 6, 2, 3, 2, 5, 4, 4, 5. It could have been the way that I rolled it that made it not turn up on the 1, or else the indents were too large and it was very heavy on the bottom

side [...] But by that it looks quite fair. At the start it looked a bit unfair, but then it looked quite fair.

*Large number of trials.* Three students trialed the dice more than 18 times and systematically recorded results to consider whether the distributions of outcomes were even. The recording strategy used involved tallying frequencies for each of the six outcomes, thus readily permitting interpretation against the hypothesis of an even distribution of outcomes:

S21: (Grade 9) [Without comment, rolls each die 20 times and records in a tally table (numerals that follow refer to the number of dots on the die): White: four 1s, nine 2s, seven 3s; Blue: one 1, three 2s, eight 3s, six 4s, one 5, one 6; Red: two 1s, three 2s, four 3s, four 4s, three 5s, four 6s] I definitely think the Blue one's weighted. You can feel it's heavier, and the White one possibly, but I don't think the Red one is at all because it's fairly even across... [Points to results] [...] The White one could be, I don't know. It could be just freaky that it came up with those 3 all the time but then again who knows.

This student only noticed the repeated numbers on the White dice after excessive prompting, as the continuation of the dialogue shows:

S21: I think the Red one is normal because it seems to have fairly even distribution with the numbers. [*I: ... the White one?*] I think it is [unfair] because it seemed to always roll 1, 2 or 3. [...] [*I: Prompts to inspect*] It doesn't seem to look like it is or anything. It could have been like a freak that it was coming up like that. [...]. [*I: Well just as it happens this one has a 2 and a 2 here...*] [...] Oh, no wonder! So it's not weighted, it's just like a cheat dice. I didn't even notice that.

Research Question 3: Association of Beliefs and Strategies

Research Question 3 concerned the association between students' beliefs and their strategies for judging fairness of dice. Table 6 shows that for the 108 students in Study 1, there was not a strong association between the levels of response for beliefs about fairness of dice and the levels of response for strategies for determining fairness, although it was statistically significant (assigning scores of 0 to 3 for the four levels,  $r = .28$ ,  $p < .005$ ). Of those with IK-Unfair or U-Fair beliefs about dice, about a quarter used IK-Idiosyncratic strategies for judging fairness, over half used

Table 6  
*Frequency of Level of Strategy by Level of Belief*

Level of strategy	Level of belief				Totals
	0 IK-Unfair	1 U-Fair	2 M-Fair qualified	3 R-Short term	
IK-Idiosyncratic	14	9	1	0	24
U-Untestable	7	6	1	0	14
M-Observational	22	32	5	4	63
R-Empirical	1	3	3	0	7
Totals	44	50	10	4	108

M-Observational strategies, and only four students performed empirical trials (R-Empirical). The 10 students who expressed M-Fair qualified beliefs generally used M-Observational or R-Empirical strategies (5 students and 3 students, respectively). Of the four students who believed in fair dice subject to short-term variation (R-Short term), none used an empirical strategy for judging fairness; all preferred observational strategies.

Of the students whose responses were used as examples earlier, S5 was one of seven students with beliefs classified as IK-Unfair and who used an U-Untestable strategy for determining fairness. S10 was an example of consistency of belief and strategy, in claiming dice are fair and then offering no strategy because dice were considered untestable. S12 was one of the three students who offered beliefs classified as Fair qualified that considered physical conditions, and who employed an empirical strategy using a small number of systematic trials.

Some responses involved wide discrepancies between level of belief and level of strategy. One student, for example, believed dice were unfair but used an empirical strategy. The belief was simply stated and the strategy used became apparent in the dialogue, which referred to classroom experience:

S22: (Grade 9) Umm, yes, they do. [*I: Which ones?*] 1, I reckon. [...] In our maths class, we did probability last year. We just rolled the dice to see what number came up. And we just wrote down how many times a certain number came up. [Rolls each 10 times, records results in tally chart for each die] This one [White] stayed in the 1 to 3 range. This one [Blue] was pretty fair, it went over all of the numbers. This one [Red] was sort of fair, but it stayed around 2 to 4 and sometimes 6.

On the other hand, one student offered a Fair qualified belief but when asked to determine fairness, she considered the dice to be Untestable. The following dialogue illustrates that despite reference to the experience of getting 1s, she believed that dice were fair because “there’s only one of each number,” a belief that overrode any systematic engagement with the actual dice in front of her. As such, her Untestable strategy led to her simple restatement of her beliefs that having one of each number was a qualification for fairness and that “repeated 1s” was an unfair outcome:

S23: (Grade 3) Sometimes. Whenever I get a 1, I always continue on getting a 1, which is sometimes a pain. [...] Yes, they all do have the same chance, because there’s only one of each number. And it is not like every side is different [...] [*I: What will you do to try and decide which ones are fair and which ones aren’t?*] Well, they are all fair, I think, because there is really no difference, because if I threw a 1, it may come up again as I said before, but it could even be something else, another number, because if I just kept on throwing 1s, that would be a funny dice, and it would probably have 1s all over it. But there’s a chance of any number coming up, really.

Overall, 43 out of 44 students who believed dice were unfair and 47 out of 50 students who believed dice were fair, suggested idiosyncratic, untestable, or observational strategies, suggesting they felt no need to test what they *knew* to be true. On the other hand, only 3 of 14 students who had a fair qualified or short-term variation view of fairness used an empirical strategy to test fairness, whereas 9 suggested weaker observational strategies, and only 2 gave idiosyncratic or untestable strategies.

## Research Question 4: Longitudinal Change

In Study 2, we consider the results for Research Question 4 in three parts related to the previous three research questions. We report longitudinal change for the 44 students interviewed twice, with respect to their beliefs about the fairness of dice, their strategies for determining fairness, and the association of these two ideas. As a group, these 44 students were judged to be representative of the original group of 108 students because the distributions of beliefs and strategies for comparable grades were similar; a statistical test was not viable because of low cell counts.

*Beliefs*

A summary of the longitudinal change of beliefs for the 44 students interviewed after 3 or 4 years for Study 2 is shown in Table 7. Nineteen students remained at the same level, and 20 responded at a higher level. Only 5 students responded at a lower level, with all responses being one level lower than their response from Study 1. As for the initial interviews in Study 1, many students responded with U-Fair beliefs, and only some qualified fairness (M-Fair qualified) or commented about short-term variation (R-Short term), possibly because there was limited motivation to do so. Whereas 20 students stated IK-Unfair beliefs in their initial interviews, only 6 did so in the later interviews. Only 1 student (S9) changed from a U-Fair belief as previously quoted to an IK-Unfair inconsistent belief in the second interview. However, in the first interview the interviewer did not ask the follow-up question about *chances* that revealed the inconsistency shown below:

S9: (Grade 7) No. [*I: You reckon they all come up about the same amount, do you?*] Yes. [*I: Do you think they all have the same chance, or do some numbers have more chance than others?*] Probably 1 and 2 have got more chance of coming up, because you don't normally score the 6s and that.

Table 7  
*Frequency of Levels of Belief of Students Interviewed in Both Study 1 and Study 2*

Level of belief (from Study 2)	Level of belief (from Study 1)				Totals
	IK-Unfair	U-Fair	M-Fair qualified	R-Short term	
IK-Unfair	5	1	0	0	6
U-Fair	9	13	3	0	25
M-Fair qualified	2	2	1	1	6
R-Short term	4	1	2	0	7
Totals	20	17	6	1	44

A sixth-grade student who stated IK-Unfair idiosyncratic beliefs in her first interview and continued to select the same numbers as unlikely in the second, provided more explanation:

- S1: (Grade 10) 3 and 4 come up most. [...] [*I: And all the others come up less often?*] Mmm. They're extremes, like 6 and 1 come up the least, and then 2 and 5 come up more frequently, then 3 and 4.

Student S4, who in the first interview believed dice to be unfair (IK-Unfair) based on experiences rolling dice with her mother, in the later interview mentioned experiences in the classroom that apparently confirmed her earlier beliefs about some numbers being more likely:

- S4: (Grade 8) I think there's other numbers that are more common than other numbers, because we did a thing in maths not long ago, when you had to I think chuck the dice 50 times, or something, and we had to record how many 5s, 4s and all that. And there's some numbers that come up more than others. Like, I think it was a 5 and a 4, they came up more.

A third-grade student's response changed from an IK-Unfair belief to a M-Fair qualified belief, noting conditions for rolling dice to avoid bias. Intonation and laughter indicated she was now parodying the idiosyncratic belief, and in the dialogue following this quote, she quickly affirmed twice that all numbers have the same chance:

- S2: (Grade 6) My side of the version? Okay, well probably with me, I have some kind of luck, okay. So I know it can be 50:50 each way of the dice, but usually if I roll it a certain way, it usually comes up a 6 and then usually.... Do you want me to tell you the way my grandma does it? She holds the dice and she talks to the dice, to say "come up a 6," "come up a 3." It doesn't work, but it works for me, okay. [Laughs] No, but usually, um no, you can't really predict which ones come up. [*I: Does it? It works for you, does it?*] Yeah, sometimes, not all the time. It all depends on the way that I throw it and what kind of table it is and all that.

One of the ninth-grade students (S8) who offered an IK-Unfair inconsistent belief in the first interview, responded in the second interview with a M-Fair qualified belief:

- S8: (Grade 12) No, I don't think so, not unless the dice is heavier on one side or something like that. But no, I don't generally think so. [*I: So the numbers all have the same chance of coming up?*] I would say so, yes.

### Strategies

A summary of the longitudinal change in strategies for the 44 students interviewed after 3 or 4 years is shown in Table 8. Twenty-six students remained at the same level, and 16 responded at a higher level. Only 2 students responded at a lower level in the second interview. Many offered IK-Idiosyncratic or M-Observational strategies in their first interview, whereas in the second interview most suggested Observational strategies.

Of the 11 students who gave IK-Idiosyncratic strategies in their first interview, in subsequent interviews 1 again offered an Idiosyncratic strategy, 1 an U-Untestable strategy, and 9 M-Observational strategies. One student (S15) appeared to misconstrue fairness in both interviews, offering IK-Idiosyncratic strategies each time:

Table 8  
*Frequency of Levels of Strategy of Students Interviewed in Both Study 1 and Study 2*

Level of strategy (Study 2)	Level of strategy (Study 1)				Totals
	IK- Idiosyncratic	U- Untestable	M- Observational	R- Empirical	
IK-Idiosyncratic	1	0	0	0	1
U-Untestable	1	2	1	0	4
M-Observational	9	3	21	1	34
R-Empirical	0	1	2	2	5
Totals	11	6	24	3	44

S15: (Grade 8) Okay, so you've got to make them so they're fair? [*I: You can have a look around them [...] how would you work out which ones are the unfair ones?*] Probably the number 6, because it's higher, and then you can get to move up more. [*I: So which ones [pointing to dice] do you think?*] Is that all on one number, then it would be fair?

One student quoted as employing an IK-Idiosyncratic strategy for judging fairness in the first interview considered possible outcomes as an M-Observational feature in the second interview, but was reluctant to use trials:

S16: (Grade 10) [Inspects each die] That one is a bit unfair [White] [grins] because they've got the same number twice. [Inspects other dice again] That one [White] is unfair. [*I: Supposing [...] someone [...] said "I think one of those is unfair." How would you convince them that they were wrong...?*] Just show them that they all had the same numbers, so they couldn't really be unfair. Because it's just chance, it's got nothing to do with anything else. [...] [*I: So what would you do to someone who comes in and says, "Look, I've just rolled this 20 times or something and it's come up really often on one of the numbers"?*] Just say it's lucky, it's got nothing to do with being unfair.

Of the 24 students who gave M-Observational strategies in their first interview, in subsequent interviews 21 again offered this kind of strategy, 1 a U-Untestable strategy, and 2 an R-Empirical strategy. One student (S17) who initially considered the M-Observational strategy of possible outcomes did so again in the second interview:

S17: (Grade 13) These [Red, Blue, and Green] each have 6 numbers whereas this [White] only has 3, so the 1, 2, and 3 have twice as much chance of appearing on this die, than any of these dice, and the 4, 5, and 6 have no chance of appearing on this, whereas they each have a 1 in 6 chance of appearing on each of these dice.

Another student who offered M-Observational strategies both times, judged the dice on weight in the first interview, and in the second interview did so again but also acknowledged the repetition of numbers on the White die:

S18: (Grade 10) [Manipulates Green die] That one feels weighted. It feels weighted on this side. [...] [Manipulates White die] That one only goes 1, 2, and 3, not other numbers. [Blue] This one feels weighted as well, except on the other side, probably for 2. [...] [*I: Is there a way you could convince me that it is weighted towards, so it would come up more on 2?*] Umm, hold it loosely on the corner, it goes down a bit, swings down a bit, like that. If you go down the other side, then it doesn't.



*Association of Beliefs and Strategies*

Table 9 shows the association of beliefs and strategies for the 44 students interviewed longitudinally in their second interviews. As for the initial interviews, there was not a strong association between the levels of response for beliefs about fairness of dice and the levels of response for strategies for determining fairness, although it was statistically significant (using the scores 0 to 3 for the levels,  $r = .29$ ,  $p < .005$ ). Compared with 3 or 4 years earlier, a higher percentage at each belief level responded with M-Observational strategies, with the highest being 84% for those with U-Fair beliefs.

Table 9  
*Frequency of Level of Strategy by Level of Belief for Study 2*

Level of strategy	Level of belief				Totals
	IK-Unfair	U-Fair	M-Fair qualified	R-Short term	
IK-Idiosyncratic	1	0	0	0	1
U-Untestable	0	3	0	1	4
M-Observational	5	21	4	4	34
R-Empirical	0	1	2	2	5
Totals	6	25	6	7	44

The student who expressed an R-Short-term variation belief but offered no testable strategy (U-Untest) was a twelfth-grade student, who was already quoted in earlier sections of this article as initially offering IK-Idiosyncratic beliefs and an U-Untestable strategy. In both interviews she made assertions that all dice are fair, although the defense was more eloquent in the second interview:

S5: (Grade 12) I think they have exactly the same chance of coming up, it just seems that the lower number comes up more often. [...] I don't know. I wouldn't know how to approach it. I'd just say that you've always got a 1 in 6 chance with any type of dice, no matter how big or how small it is. You've always got a 1 in 6 chance because usually each side is equal. It has to be equal to make a dice. So it wouldn't matter how many times you tested it, there'd always be a 1 in 6 chance of getting the number that you're after.

One student whose responses in the first interview have already been quoted, involving a M-Fair qualified belief and an R-Empirical strategy, in the second interview added a notion of R-Short term and increased the number of trials for each die from 12 to 20. Her recording strategy also improved, from a column listing the sequences of outcomes for each die, to a structured table for each die with a column labeled "No" with the values 1 to 6, a column labeled "Tally" to record the outcomes, and a column labeled "f" to denote the frequency as the sum of the tally marks after 20 trials:

S12: (Grade 10) No. [*I: No?*] In theory they don't. [*I: In practice?*] I think that because dice are rolled so many times, it's probably relatively even, so no. Unless it's a biased dice and it has a weight in it. [...] [Picks up pen to write on paper] What I think I'll do is.... How much time do I have? [...] I think I'll roll each one 20 times. And I'll do a chart for each one.... A frequency table. [Rolls Green 20 times, frequencies for the outcomes 1, 2, 3, 4, 5, 6 were 5, 7, 3, 2, 0 and 4, respectively] Okay, I think there might be a weight down there [Points to corner of die] seeing that there are so many 1s and 2s and 6s. But I'm not sure. Might just be a coincidence. [...] Rolls Blue and Red dice....] [Rolls White die 17 times, frequencies for the outcomes 1, 2, 3 were 6, 7, and 4, respectively] I won't go on, because I can see this is very obviously biased towards 1, 2, and 3, mainly 2, but it's probably the weight around somewhere there [face with 3]. So really I don't think um, I think probably most of them were biased except for the Blue one.

## DISCUSSION

Implications of the results are discussed with respect to beliefs about fairness of dice, strategies for assessing fairness, and conceptual associations and development. We also include suggestions for future research and implications for teaching.

### *Beliefs About Fairness of Dice*

Beliefs about fairness of dice have been explored by various researchers, with “non-statistical” beliefs observed to be held by young children (J. Truran, 1985; K. Truran, 1995), middle school students (Amir & Williams, 1999), and even adults (Konold et al., 1993). The findings of the study reported here confirm that many students hold idiosyncratic and contradictory beliefs throughout the years of schooling. The percents of students in Study 1 believing in fairness—38% of third graders, 60% of fifth to seventh graders, 79% of ninth graders—were similar to those of large scale survey-based studies (Green, 1983; Kerslake, 1974; Watson et al., 1997). The framework for describing responses observed in this study reflected to some extent our previous research experience about students' understanding of concepts about chance and data and those from a more general developmental model (Biggs & Collis, 1982, 1991). Ikonc-Unfair beliefs involved personal experience, storytelling, and lack of recognition of contradictions. Unistructural-Fair beliefs were expressed as single statements, without amplification; Multistructural-Fair qualified beliefs were more sophisticated in combining relevant features of dice that could affect fairness; and Relational-Short term variation beliefs resolved the conflict between experiences of short-term “unfair” outcomes and long-term “fair” chances.

A potential limitation of the study reported here is that the protocol did not specifically ask for the reflections evident in Multistructural-Fair and Relational-Short term levels. Hence some students, for example those responding at the Unistructural-Fair level, may have been able to make comments at the Multistructural-Fair or Relational-Short term levels but did not realize it was relevant. Students who indicated that dice are unfair were asked to explain which

numbers come up more often. Students who indicated dice are fair were asked to clarify if this meant all numbers have the same chance. In some cases (e.g., S11), this simple request for clarification elicited a higher-level response. Additional probing for explaining beliefs at higher levels was avoided, however, because we did not want to influence responses to the second part of the protocol by introducing extra concepts if the student did not suggest them.

Using the four-level framework, 5 students (11%) responded at a lower level 3 or 4 years later, however, only 1 changed from a Unistructural-Fair belief to an Ikonic-Unfair belief. Of those who could improve, that is, from the Ikonic-Unfair, Unistructural-Fair, and Multistructural-Fair levels to a higher level, of 43 students, 20 students did and were mostly from the Ikonic-Unfair category (15 students). The outcomes of our study suggest that over the years between interviews, the majority of students took on the view that all numbers on dice have the same chance of coming up when tossed. Whether this arose from performing trials in classrooms, from accepting teachers' statements as authority, or from other personal experience is unknown in most cases. What is known from some student comments (e.g., S4 and S22), however, is that even classroom trialing can be interpreted or remembered in such a way as to fail to engender belief in fairness, and perhaps even to reinforce beliefs that dice are unfair.

Other researchers (Amir & Williams, 1999; J. Truran, 1985; K. Truran, 1995) have observed that some students have beliefs that God or their mental powers determine some outcomes in chance situations. In our study, God was not mentioned by any student in relation to determining the outcomes of dice rolling. The cultural influences appeared to be more likely related to experiences playing games where the conflict of observed outcomes versus parental assurance that "all dice are fair" may be a factor. In considering the contradictory comments made by some students in explaining their beliefs (e.g., "Some numbers come up more often, but all dice are fair."), however, there was some indication of a subtle distinction between frequencies and chances, most often where students believed some numbers occur more often but all have the same chance. Such inconsistencies are reminiscent of those found by Konold et al. (1993). It may be that a question about "chances" is associated with the future, or as being more mathematical or theoretical, whereas a question about frequency of occurrence of numbers is associated with personal experience in the past and hence factual evidence. It could also be that in this case students were simply agreeing to both forms of the question asked due to lack of understanding and a desire to please the interviewer.

### *Strategies for Assessing Fairness of Dice*

In contrast to the research done on students' beliefs about the fairness of dice, there has been very little research on students' naive strategies for determining fairness, except for the preliminary analyses related to this study (Lidster et al., 1995, 1996; Watson & Chick, 2001). The framework for describing students' strategies reflected both the structural complexity, as suggested by the cognitive

model of Biggs and Collis (1982, 1991), and the statistical appropriateness of the content of responses. Students using Ikonc-Idiosyncratic strategies were not able to engage with the notion of fairness implied by the task. At the Unistructural-Untestable level, a response associated with the assertion that “all dice are fair” and thus untestable was considered an inappropriate strategy for the *task of testing fairness*. At the Multistructural-Observational level, students tended to use observational strategies, which one might consider as a partial response or even optimal in considering the White die with repeated numbers in reference to the other dice. These strategies were not optimal however in a statistical sense where hypotheses need to be tested. It may be that some students were satisfied with what might be considered a functional response that addressed the task, whereas others may never have been exposed to classroom experiences reinforcing the necessity to test their beliefs. At the Relational-Empirical level, strategies reflected an integrated view of a statistical trial to confirm the nature of the dice’s behaviors. These were both structurally complex and statistically the most appropriate. It is interesting to note that few students using empirical strategies also employed observational strategies, suggesting that the categorization of one dominant strategy was appropriate for the analysis. The distinction noted between students who performed few or a larger number of trials reflects a similar split that we noticed for students considering sampling issues (Watson & Moritz, 2000a), where sample size was considered a major threshold in the development of understanding of sampling. A higher proportion of students in that study suggested large samples, but we did not ask students to collect the data and hence time was not an issue. The tendency for students not to respond to repeated prompts to perform trials may be associated with lack of exposure in the classroom and points to recommendations made in a later section.

### *Conceptual Associations and Development*

Both beliefs about fairness and strategies for judging fairness were examined in this study as they are important components of school curricula (AEC, 1991; NCTM, 2000) for understanding theoretical probability based on reasoning about equally likely outcomes, and assessing probability based on relative frequency of empirical outcomes. Curricular recommendations link belief and strategies when they include empirical trialing in the classroom in order to test or confirm beliefs in fairness. Kerslake (1974) questioned whether empirical trialing would convince students of the fairness of dice, and little research evidence has previously addressed this issue. In the study reported here, the similarities between levels of beliefs (see Table 2) with respect to structure of responses (idiosyncratic, single aspect, multiple aspects, relational) and levels of strategies (see Table 3) with respect to the aspects of focus (idiosyncratic, simple beliefs, observation of physical characteristics, variation of empirical outcomes), offer some structural indication that beliefs and strategy might be associated, and that development in one might be associated with development in the other.

The response data from this study, however, gave little evidence to support the existence of an association between beliefs about fairness and strategies for judging fairness. In fact, there was some evidence to the contrary when the majority of students with the highest level of beliefs and who acknowledged awareness of variation in short-term outcomes used observational strategies in judging fairness. On one hand, these responses may have been based on an appreciation that small-scale trialing would not yield a definitive judgment about fairness of dice. On the other hand, one ninth-grade student (S22) conducted trials and drew conclusions that reflect the type of experience that might be expected in the classroom, but the initial belief that 1 comes up more often was not consistent with these later comments. This response may be an example of the inconsistency between out-of-school experience or intuition and in-school reasoning or "knowledge." J. Truran (1985) asked a tenth-grade student if some numbers were easier to obtain when tossing a die. The student responded, "Learning what we did in maths, no, not really. When you do it [...] the middle numbers are easiest [...] [teachers] tell you one thing, but when you go home and do it, it doesn't seem to be the same" (p. 73). These observations also reflect the experiences described by the mathematics major mentioned in the introduction to this article. The lack of association between beliefs and strategies may indicate that for some students beliefs involve a classical approach to probability-based assumptions of equally likely outcomes, which are quite divorced from the empirical approach of judging probability based on long-term relative frequency (Borovcnik & Bentz, 1991).

We found that levels of beliefs and strategies were significantly correlated with SOLO response levels to tasks involving average (Watson & Moritz, 2000b) and comparing data sets (Watson & Moritz, 1999),  $.32 \leq r \leq .54$ ,  $p < .005$  in all cases, more strongly than they were with each other (for the original interviews from Study 1,  $r = .28$ ,  $p < .005$ ). For the concept of average, development was displayed in relation to building up a representational idea of average in four levels beginning with idiosyncratic ideas, and then applying this in problem-solving settings, for example, with a weighted mean. For the comparison of two data sets, development was displayed in three levels of increasingly complex structure of reasoning for data sets of equal size and then in three higher, but similarly structured levels, for data sets of unequal size. Lack of fairness of the comparison was a consideration for the unequal sized sets and at the top two levels of response the arithmetic mean was likely to be employed to provide a fair comparison. Although not strong, these associations among varied tasks may indicate that performance on these tasks develops, to some degree, hand in hand, perhaps due to the underlying notion of representativeness (Kahneman & Tversky, 1972) that is common for ideas of fairness and average in a variety of contexts.

The fact that there was a tendency for students to respond at higher levels with increasing grade in the initial interviews and to improve over a 3- or 4-year interval (45% of students for beliefs and 36% for strategies) is in contrast with findings of other researchers. Batanero and Serrano (1999) for example found consistent performance for 14- and 17-year-olds on survey items related to randomness, and

Fischbein and Schnarch (1997) found that performance for some probabilistic intuitions decreased with age. These other researchers, however, used survey items and asked for explicit decisions instead of focusing on open-ended questions about beliefs and strategies. The relationship between these two types of inquiry—based on surveys or interviews—and what they highlight about student understanding is an area for future research. The improvement in levels of response in this study was not as great over 3 or 4 years as that observed in relation to average (Watson & Moritz, 2000b) or comparing two data sets (Watson & Moritz, 1999). Positive change did occur for many of those who initially offered beliefs or strategies at the ikonic level. Few students, however, changed from simple Unistructural-Fair beliefs and Multistructural-Observational strategies, possibly because students felt these types of response minimally satisfied the question and there was no need to elaborate. We also observed this phenomenon of minimal responding for some students representing and interpreting pictographs (Watson & Moritz, 2001a). For understanding of average, 73% improved ( $n = 43$ ) with four students at the top level each time, whereas for comparing two data sets, 62% improved ( $n = 42$ ) with two students at the top level each time. This difference in improvement percents is likely to be associated with specific attention to average, with some spin-off for comparing data sets, in the middle years of schooling, and the lack of attention to carrying out trials with random generators.

### *Future Research*

Several avenues exist for future research following the outcomes observed in this study. As noted, there was no educational intervention as part of the design of the project in the years between the interviews. We do not know whether or not teachers engaged students in discussion related to fairness and gathering empirical evidence in the intervening time. In either case, there was little evidence of students developing increased motivation or ability to gather empirical data with an awareness of the need for large sample size because of short-term variation of outcomes. This was true even at higher grade levels (e.g., Grade 12) that were not included in the initial interviews. One variation on the approach used in this study would be to include a teaching component. Various teaching methodologies could be used in the classroom, including collaborative groups designing and carrying out trials, individual project work, or demonstrations by the teacher, to see which is the most effective. Both short- and long-term retention could be monitored. The content of the teaching could address the issues of belief, judgment of fairness, trialing procedures, and the relationship among them. This could be implemented in the classroom with one random generator, say coins or spinners, and tested in pre- and post-learning interviews with another random generator, say dice. Transfer would be expected if students can generalize the idea of random generator, however, it is not a certainty if students hold beliefs specific to each random generator (J. Truran, 1985).

A variation on the interview protocol would be to allow more time and at some point, depending on the students' own suggestions, direct them to perform trials.

The researchers could then take note of the method of rolling, the number of trials, and the method of recording. On the other hand, these factors could be controlled and the interest placed in the conclusions drawn from the data set collected, in the light of the pattern and variation displayed in the distribution of outcomes. Other random generators, such as spinners, coins, or objects drawn from opaque containers, could be used, particularly to determine if beliefs about dice are more idiosyncratic than beliefs about other such devices. It is also possible in the interview setting to create cognitive conflict, say from responses of other students who were videotaped earlier, and see to what extent these responses influence interviewees' beliefs. We did this in another study (Watson & Moritz, 2001b) in relation to a proportional probability problem drawing marbles from containers, and often the prompting with cognitive conflict led to improved responses.

Although we acknowledge that the observation of increased structural complexity as reflected in the SOLO model may not be as obvious in this study as in the contexts of questioning about average (Watson & Moritz, 2000b) or comparing two data sets (Watson & Moritz, 1999), the indication of using imaginative, single, multiple, and related elements is present. The fact that for beliefs these higher-level responses were not explicitly solicited provides a realistic description of students' off-the-cuff thinking. The small number of such responses may indicate that classroom experiences should encourage more self-exploration of beliefs by students at all grade levels. Future research may, through different forms of questioning, refine these levels of response but they provide a theoretically based starting point.

### *Teaching Implications*

Several students' responses from this study cause concern in terms of what may be happening in the classroom. Consider, for example, the eighth-grade student (S4) who commented about her classroom trials throwing a die 50 times, which led her to conclude 5 occurred more often. Perhaps of even more concern was the response provided by the tenth-grade girl (S1) who suggested a peaked distribution for tossing a single die. This distribution was also preferred by 36% of students to describe the likely outcome of 60 trials of a fair die in the study by Green (1983). The idea that the middle numbers are more likely may be related to experiences of tossing two fair dice and summing the outcomes, and then transferring the shape of the distribution to tossing a single die. We believe students need experiences with dice in schools and considerable time to reflect on these experiences, guided by the teacher—not just in complex situations (e.g., Bright, Harvey, & Wheeler, 1981; Maher, 1998; Vidakovic, Berenson, & Brandsma, 1998) but in the simple context of a single die (Watson, 2002). Emphasis also needs to be placed on the variation in outcomes that one would expect in the short run.

The inconsistencies between beliefs about and strategies for determining fairness point to dilemmas for teachers even when they directly address the fundamental issues. If students do not believe dice are fair to start with, will a lesson trialing real dice help them to change their beliefs or not? Perhaps students "learn" to repeat

what the teachers want them to say—that dice are fair. This is the “in-school” belief. What they believe outside the mathematics classroom may be another matter. We believe this is a point requiring up-front discussion in the classroom (Burrill & Romberg, 1998), perhaps even including examples from the media where claims of luck or foresight are made (Watson, 1993). In the overall context of the probability part of the mathematics curriculum, the topics addressed here provide an excellent opportunity to address all three perspectives on probability: subjective, frequency, and theoretical (Borovcnik & Bentz, 1991; Metz, 1997). If done from the beginning, perhaps students will develop a more balanced and circumspect belief structure. Further suggestions by Horvath and Lehrer (1998) and Metz (1998) include structural aspects that are likely to assist elementary children in building appropriate statistical approaches to judging fairness related to probability. There is also the possibility to compare and contrast fairness in probability with fairness in sampling (Jacobs, 1999; Lajoie, Jacobs, & Lavigne, 1995). This relationship becomes an important issue in the senior years of schooling.

The outcomes of our study have convinced us of the importance of two instructional strategies when introducing random generators to students, at whatever grade the introduction is made. One is the use of concrete materials. Computer simulations of dice outcomes were not included as part of the protocol in this study, and it is our view that only students at the high levels of trialing, or who can be prompted to these levels, can be expected to benefit from computer simulation. The abstract nature of simulation is far removed from the hands-on approach using physical objects as a starting point to build understanding about outcomes of tossing dice. This is not to say that handling dice is without its difficulties. Our observations were that many students found it very difficult to roll dice in an unbiased fashion. We realized that this was not a trivial exercise when we ourselves conducted trials of the loaded dice. The other instructional strategy is the focused use of language, particularly related to the word “chance” itself. One aspect of this has been mentioned in relation to the outcomes for students’ beliefs. Specific discussion should take place about frequencies as observed phenomena and chances as theoretical entities. It would appear that the expectation of frequencies over increased numbers of trials approaching theoretical chances (probabilities) needs much explicit discussion. There is also the aspect of “chance” as the mechanism that provides the variation from theoretical expectations. One student (S16), for example, in explaining the lack of need to test her belief in fairness, attributed deviations to “lucky chance.” Although random behavior is an appropriate alternative description, the colloquial use of terms like “chance” in different contexts means that teachers and researchers must always clarify their use of terminology.

Although curriculum documents and textbooks assume that dice are fair as a starting point for work in probability, Peterson’s comment (1998) about the likely bias of “store-bought” dice suggests that perhaps this is not a good assumption to make. In the days when only theoretical probability based on sample spaces was taught, with no experimentation in the classroom, this assumption was adequate and fit the model of the mathematics curriculum. Today, however, it may be more



appropriate to put *questioning the fairness of the random generator* high on the curriculum agenda. Students' initial contact with probability can involve links to other parts of the chance and data curriculum: Performing trials and using them to make decisions to justify the physical model they are going to use. Having developed empirical techniques for handling data and confidence that dice are fair based on the results of trials using single random generators, students can then transfer these strategies and beliefs in modeling and describing compound events, such as the sum of two dice, and more complex phenomena, such as the fairness of games.

## REFERENCES

- Amir, G. S., & Williams, J. S. (1999). Cultural influences on children's probabilistic thinking. *Journal of Mathematical Behavior*, 18, 85–107.
- Australian Education Council. (1991). *A national statement on mathematics for Australian schools*. Carlton, Australia: Author.
- Australian Education Council. (1994). *Mathematics - work samples*. Carlton, Australia: Author.
- Batanero, C., & Serrano, L. (1999). The meaning of randomness for secondary school students. *Journal for Research in Mathematics Education*, 30, 558–567.
- Bennett, D. J. (1998). *Randomness*. Cambridge, MA: Harvard University Press.
- Bernstein, P. L. (1998). *Against the odds: The remarkable story of risk*. New York: John Wiley and Sons.
- Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy (structure of the observed learning outcome)*. New York: Academic Press.
- Biggs, J. B., & Collis, K. F. (1991). Multimodal learning and the quality of intelligent behaviour. In H. A. H. Rowe (Ed.), *Intelligence: Reconceptualisation and measurement* (pp. 57–76). Hillsdale, NJ: Lawrence Erlbaum Associates.
- 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 Academic Publishers.
- Bright, G. W., Harvey, J. G., & Wheeler, M. M. (1981). Fair games, unfair games. In A. P. Shulte (Ed.), *Teaching Statistics and Probability* (pp. 49–59). Reston, VA: National Council of Teachers of Mathematics.
- Burrill, G., & Romberg, T. A. (1998). Statistics and probability for the middle grades: Examples from *Mathematics in Context*. In S. P. Lajoie (Ed.), *Reflections on statistics: Learning, teaching and assessment in grades K–12* (pp. 33–59). Mahwah, NJ: Lawrence Erlbaum Associates.
- Fischbein, E., & Gazit, A. (1984). Does the teaching of probability improve probabilistic intuitions? An exploratory research study. *Educational Studies in Mathematics*, 15, 1–24.
- Fischbein, E., & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research in Mathematics Education*, 28, 96–105.
- Garfield, J., & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: Implications for research. *Journal for Research in Mathematics Education*, 19, 44–63.
- Green, D. R. (1983). Shaking a six. *Mathematics in Schools*, 12(5), 29–32.
- Green, D. R. (1986). Children's understanding of randomness: Report of a survey of 1600 children aged 7–11 years. In R. Davidson & J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics* (pp. 287–291). Victoria, BC: The Organizing Committee, ICOTS2.
- Green, D. (1991). A longitudinal study of pupils' probability concepts. In D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics. Volume 1: School and general issues* (pp. 320–328). Voorburg, The Netherlands: International Statistical Institute.
- Green, D. (1993). Data analysis: What research do we need? In L. Pereira-Mendoza (Ed.), *Introducing data analysis: Who should teach it and how?* (pp. 219–239). Voorburg, The Netherlands: The International Statistical Institute.

- Horvath, J. K., & Lehrer, R. (1998). A model-based perspective on the development of children's understanding of chance and uncertainty. In S. P. Lajoie (Ed.), *Reflections on statistics: Learning, teaching and assessment in grades K–12* (pp. 121–148). Mahwah, NJ: Lawrence Erlbaum Associates.
- Jacobs, V. R. (1999). How do students think about statistical sampling before instruction? *Mathematics in the Middle School*, 5, 240–263.
- Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3, 430–454.
- Kerslake, D. (1974). Some children's views on probability. *Mathematics in School*, 3(4), 22.
- Konold, C., Pollatsek, A., Well, A., Lohmeier, J., & Lipson, A. (1993). Inconsistencies in probabilistic reasoning of novices. *Journal for Research in Mathematics Education*, 24, 392–414.
- Lajoie, S. P., Jacobs, V. R., & Lavigne, N. C. (1995). Empowering children in the use of statistics. *Journal of Mathematical Behavior*, 14, 401–423.
- Lidster, S. T., Pereira-Mendoza, L., Watson, J. M., & Collis, K. F. (1995, November). *What's Fair for Grade 6?* Paper presented at the Annual Conference of the Australian Association for Research in Education, Hobart, Tasmania.
- Lidster, S. T., Watson, J. M., Collis, K. F., & Pereira-Mendoza, L. (1996). The relationship of the concept of fair to the construction of probabilistic understanding. In P. C. Clarkson (Ed.), *Technology in mathematics education: Proceedings of the Nineteenth Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 352–359). Melbourne: Mathematics Education Research Group of Australasia.
- Maher, C. (1998). Is this game fair? The emergence of statistical reasoning in young children. In L. Pereira-Mendoza, L. S. Kea, T. W. Kee, & W. Wong (Eds.), *Statistical education: Expanding the network: Proceedings of the Fifth International Conference on Teaching of Statistics* (pp. 53–59). Singapore: International Association for Statistical Education.
- 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, The Netherlands: IOS Press.
- Metz, K. E. (1998). Emergent ideas of chance and probability in primary-grade children. In S. P. Lajoie (Ed.), *Reflections on statistics: Learning, teaching and assessment in grades K–12* (pp. 149–174). Mahwah, NJ: Lawrence Erlbaum Associates.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.
- Ministry of Education. (1992). *Mathematics in the New Zealand curriculum*. Wellington, New Zealand: Author.
- Moore, D. S. (1990). Uncertainty. In L. A. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 95–137). Washington, D.C.: National Academy Press.
- National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.
- Peterson, I. (1998). *The jungle of randomness: Mathematics at the edge of uncertainty*. London: Penguin.
- Ritson, R. (2000). A question of choice. *Australian Primary Mathematics Classroom*, 5(3), 10–14.
- 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: National Council of Teachers of Mathematics and MacMillan.
- Streefland, L. (1991). *Fractions in realistic mathematics education: A paradigm of developmental research*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Truran, J. (1985). Children's understanding of symmetry. *Teaching Statistics*, 7(3), 69–74.
- Truran, K. (1995). Animism: A view of probability behaviour. In B. Atweh & S. Flavel (Eds.), *Galtha: Proceedings of the Eighteenth Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 537–542). Darwin, Australia: Mathematics Education Research Group of Australasia.

- Vidakovic, D., Berenson, S., & Brandsma, J. (1998). Children's intuition of probabilistic concepts emerging from fair play. In L. Pereira-Mendoza, L. S. Kea, T. W. Kee, & W. Wong (Eds.), *Statistical education: Expanding the network: Proceedings of the Fifth International Conference on Teaching of Statistics* (pp. 67–73). Singapore: International Association for Statistical Education.
- Watson, J. M. (1993). Introducing the language of probability through the media. In M. Stephens, A. Wayward, D. Clarke, & J. Izard (Eds.), *Communicating mathematics - Perspectives from current research and classroom practice in Australia* (pp. 119–139). Melbourne: Australian Council for Educational Research.
- Watson, J. M. (1994). Instruments to assess statistical concepts in the school curriculum. In National Organizing Committee (Ed.), *Proceedings of the Fourth International Conference on Teaching Statistics. Volume 1* (pp. 73–80). Rabat, Morocco: National Institute of Statistics and Applied Economics.
- Watson, J. M. (2001). Longitudinal development of inferential reasoning by school students. *Educational Studies in Mathematics*, 47, 337–372.
- Watson, J. M. (2002). When  $2 + 2 \neq 4$  and  $6 + 6 \neq 12$  in data and chance. *New England Mathematics Journal*, 34(2), 56–68.
- Watson, J. M., & Chick, H. L. (2001). Factors influencing the outcomes of collaborative mathematical problem solving: An introduction. *Mathematical Thinking and Learning*, 3 (2 & 3), 125–173.
- 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. *Mathematics Education Research Journal*, 10, 103–127.
- Watson, J. M., & Moritz, J. B. (1999). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics*, 37, 145–168.
- Watson, J. M., & Moritz, J. B. (2000a). Developing concepts of sampling. *Journal for Research in Mathematics Education*, 31, 44–70.
- Watson, J. M., & Moritz, J. B. (2000b). The longitudinal development of understanding of average. *Mathematical Thinking and Learning*, 2 (1 & 2), 11–50.
- Watson, J. M., & Moritz, J. B. (2001a). Development of reasoning associated with pictographs: Representing, interpreting, and predicting. *Educational Studies in Mathematics*, 48, 47–81.
- Watson, J. M., & Moritz, J. B. (2001b). The role of cognitive conflict in developing students' understanding of chance measurement. In J. Bobis, B. Perry, & M. Mitchelmore (Eds.), *Numeracy and beyond: Proceedings of the Twenty-Fourth Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 523–530). Sydney: Mathematics Education Research Group of Australasia.

## Authors

---

**Jane M. Watson**, Faculty of Education, University of Tasmania, Private Bag 66, Hobart, Tasmania 7001, Australia; Jane.Watson@utas.edu.au

**Jonathan B. Moritz**, Faculty of Education, University of Tasmania, Private Bag 66, Hobart, Tasmania 7001, Australia; Jonathan.Moritz@utas.edu.au