THE EFFECTS OF INSTRUCTION ON LIKELIHOOD MISCONCEPTIONS

Paul Ayres
University of New South Wales
Australia

Jenni Way
University of Western Sydney
Australia

This paper reports an investigation into the effects of instruction in probability concepts on the decision making strategies of twenty-four 11-12 year olds. The instruction, based on small-group practical activities, had an overall positive influence on performance in specific probability tasks. It was also found that the particular experiences within the small groups of students had a strong influence on decision-making strategies in the final 'test' tasks. Groups that experienced sets of random outcomes in their activities that were not representative of the structure of the sample space tended to use inappropriate reasoning in later tasks.

Research into probabilistic reasoning has identified various strategies used by people in situations including sequences of randomly generated outcomes, and in particular involving the expectation (or prediction) of the most likely 'next' outcome. One such strategy that has received considerable attention is *representativeness* (for example: Fischbein & Schnarch, 1997; Shaughnessy, 1981), which is the expectation that a random set of outcomes should be representative of the composition of the sample space. Amir, Linchevski & Shefet (1999), working with 11 and 12 year olds explained that;

> The 'representativeness' heuristic includes two distinct and independent dimensions: the tendency to expect a sample space to reflect the numerical proportion of the parent population; the tendency to expect the sample not to be too orderly, to look 'random'. (p. 2-32)

Closely related to *representativeness* is the type of thinking known as *negative recency* (Fischbein & Schnarch, 1997) where there exists the expectation that as the frequency of a particular outcome increases the probability of that outcome occurring again decreases. For example, when repeatedly flipping a coin, a run of heads would lead to the expectation of the next flip being a tail.

Another, little studied, influence on decision-making is the confirmation or refutation of the 'prediction' by the actual next outcome. Truran (1996), working with a *known sample space*, analysed the changes in prediction of primary and secondary students in regards to the next outcome. One finding was that when the more-likely outcome was predicted, it didn't really matter whether the next outcome confirmed or refuted that prediction. However, if a less-likely outcome was predicted, the subject was highly likely to change the prediction, particularly if the following outcome refuted the less-likely prediction. Similarly, Ayres & Way (1998, 1999), working with *unknown sample spaces*, found evidence that upper primary-aged students would change their prediction patterns according to how successful they were in their predictions. Although students would choose the most frequently occurring outcome under specific conditions, they would change strategy if their predictions were not rewarded. Consequently, Ayres & Way (1999, 2000) argued that children may be influenced in their probability judgments by confirmation or refutation of their predictions rather than by their knowledge of the overall situation.

In the context of mathematics education, the apparent instability of students' understanding of basic probability concepts such as randomness and sample space, makes the influence of instruction an important area for investigation. Watson & Moritz (1998) reported that after five years inclusion of probability and statistics in an Australian state's curriculum (primary and lower secondary), no improvement in students' performance in chance measurement tasks was found. However, Jones, Thornton, Langrall & Mogill (1996, 1999) found clear indications of improvement due to instruction in chance concepts in Grade 3 children. The Jones et al. (1999) study suggested several key aspects of instruction; overcoming misconceptions about sample space and its relationship to likelihood, the application of part-part and part-whole reasoning, and the development of language to describe probabilities. Further research into significant experiences that may help or hinder the development of understanding of probability concepts is required. For example, little is known about the impact of commonly used classroom-teaching techniques, such as group practical work and whole class discussions, on the development of probability ideas.

The study reported here is the latest in a series of investigations into probabilistic reasoning and utilises the established 'controlled randomness' video (see Ayres and Way, 1999) as an assessment instrument . The aim of this particular study was to carry out an initial investigation of the effect of instruction, based around small-group practical activities, on the basic decision-making strategies of upper-primary students.

INSTRUCTION PERIOD

Participants. Twenty-four grade-six students (11-12 year-olds) from an independent Sydney (Australia) primary school for girls participated in this study. No student had been formally taught probability theory in any mathematics classes. The students were drawn from the top mathematics class of the grade and a state-wide numeracy test indicated that each participant ranked above the State average on this test. The students were randomly assigned to six groups

Procedure and materials. The instruction period consisted of two 1-hour sessions over consecutive days. Because of space restrictions, this paper will focus on the first session which consisted mainly of a small group practical activity. Initially (approximately ten minutes), a whole-class discussion with one of the researchers (Investigator 1) was completed on the basic ideas associated with likelihood and chance. Although the class had received no previous school-based instruction on likelihood, it was clear from the discussion that many students had a reasonable understanding of likely and unlikely events in the real world. Furthermore, some students demonstrated knowledge of theoretical probabilities associated with obtaining a head in coin-flipping ("1 in 2", "fifty-fifty") or a six in dice-rolling ("1 in 6").

Each group received a paper bag containing ten coloured tiles. Two groups received a bag containing 5 Green, 3 Yellow and 2 Blue tiles; two groups received a bag with 6G, 3Y and 1B tiles; and the remaining two groups received a bag with 7G, 2Y and 1B tiles. Ratios varied to investigate the impact that these differences might have on the overall results. Prior to the start of this phase, a demonstration of a game was given by the researcher to the whole class using a bag with colours in the ratio 1: 3: 6. A total of four games were completed. The only information given to the students was that the bags contained green, yellow and blue tiles. Students were required to predict the colour of a tile before it was withdrawn from the bag. Students took turns within groups to select a tile, which was returned to the bag before the next selection. Students were told that it was a game and the winner would be the one with the most correct predictions overall. Students were required to record each prediction and the actual colour that occurred in a booklet. A game consisted of five predictions. After each game, students were required to tally their correct predictions and discover the winner(s) for that game in their group. Furthermore, after each game, students were asked to discuss the game within their groups and to ascertain why the *winner* won, and to record this reason/discussion in their booklets.

On finishing the last game, students were asked to count the number of correct predictions they made over the four games and to discover who the overall winner was in their group. They were then asked to reflect on winning strategies and how they could have improved their own predictions. After all tasks were completed, students were allowed to examine the contents of the bag. This session was closed with a whole class discussion around the main idea that there were more greens in the bags, therefore the best prediction strategy was to choose green.

The second session (which will be reported in more detail at a later date), focused on two central themes related to the structure of the sample space and its relationship to likelihood. Firstly, selecting the most frequently occurring colour in the outcomes was a good strategy. Secondly, this strategy does not always work as less likely events can occur. A practical activity based on the strategy of choosing the most frequently occurring colour over a number of trials was completed.

Results. The mean number of greens (there were more greens in each bag than any other colour) predicted by each group is reported in Table 1. There was a significant difference between the groups on this statistic, measured by a 1-way ANOVA; $F(5.18) = 5.43$, $p < 0.01$. In terms of the number of correct predictions made (see Table 1), a 1-way ANOVA also revealed a significant difference between groups; $F(5,18) = 3.67$, $p < 0.05$. Whereas these results are not necessarily surprising considering the differing bag contents; it is clear that Groups 2 and 3 observed a reduced frequency for green to what might be expected from the theoretical probabilities: 35% compared with 50% for Group 2, and 40% compared with 60% for Group 3. The small number of occurrences of green influenced both groups' selection of green and ultimately their prediction successes, as there was a correlation of 0.89 (Pearson product-moment coefficient, $p < 0.001$) between prediction success and choice of green. Students who regularly chose green were more successful in their predictions. The written responses from individuals indicated that students from Groups 1,4, 5 and 6 believed that green was the dominant colour in the sample space and students would enhance their prediction rates if they chose more green. In contrast, Group 2 concluded that luck was involved and success depended upon being able to “spot the patterns”. Group 3 generally believed that there were more yellows in the bag and success depended upon “knowledge of previous games”.

Table 1: *Quantitative Group data for instructional phase*

Group	Ratio of G:Y:B in Bag	Actual colour (G:Y:B) outcomes over 20 trials	Mean number of greens predicted over 20 trials	Mean number of correct predictions over 20 trials
1	5: 3: 2	12: 5: 3	9.8	8.3
2	5: 3: 2	7: 7: 6	7.3	6.8
3	6: 3: 1	8: 8: 4	6.5	5.3
4	6: 3: 1	12: 7:1	11.8	9.5
5	7: 2: 1	15: 3: 2	13.5	11.5
6	7: 2: 1	13: 4: 3	11.8	9.8

Whereas, many students (particularly in groups 1, 4, 5 and 6) were able to give reasons consistent with an understanding of likelihood, the winner of Group 2 (9 out of 20) gave a very unexpected answer during an interview.

> I worked out a theory. The teacher (researcher) is English and he pulled out a yellow tile. My dad’s English and I also pulled out a yellow tile. Alison’s dad is Australian and Australia is on the opposite side of the world to England, therefore she would pull out a blue tile and she did. Maria’s dad is Greek, therefore she should pull out a green tile and she did.

In Group 2, students were taking turns selecting from the bag. The winner changed her choice according to who was making a selection. By coincidence, she was the most successful of her group, and this misconception became her “successful” strategy. Overall, quantitative and qualitative data revealed that most students demonstrated a good understanding of likelihood in this domain. However, it became evident that selection strategies and the reasons given by individuals tended to converge within the groups.

TEST PERIOD

Participants. The 24 participants were the same students who completed the instruction period. However, two students (one each from Groups 1 and 3) were absent.

Procedure and materials. To test the effectiveness of the instructional phase, the students were given prediction tasks based around a video recording. The video, previously constructed by Ayres and Way (1998, 1999) with pre-ordained outcomes, featured a presenter making thirty selections of coloured balls from a box with replacement. Students were required to predict the next colour after observing the five previous selections. In all, six predictions were required. In the video, 19 whites (63%), 7 blues (23%) and 4 yellows (13%) were drawn from the box. However, the emerging colour sequence was manipulated so that the less likely outcomes (blue and yellow) appeared consistently (four out of five) at the prediction locations, whereas the most likely colour (white) appeared only once. As a consequence of this design, students found that predicting a number of whites was not a successful strategy and switched from using "the most likely strategy" to strategies based on misconceptions such as colour patterns and negative recency (see Ayres and Way 1999, 2000 for more detail). In this present study, it was anticipated that this particular task would prove a considerable test of student beliefs in employing the "most likely" strategy consistently without reverting to common misconceptions.

Previous research by Ayres and Way (1999, 2000) also found that knowing or not-knowing the sample space made no overall difference to the prediction strategies employed on this video task. This aspect was also investigated in this study by randomly assigning students to two groups. One group (*sample space known*) was informed that the box contained 10 balls (6 white, 3 blue and 1 yellow), which was approximately equivalent to the experimental probability of the sequences. In contrast, the second group (*sample space unknown*) were only given information about the colours of the balls (some white, blue and yellow balls in the box). Both groups observed the same video recording. After each prediction, students were required to give reasons for each decision. All students were told that the prediction tasks were a game and they should try to predict as many correct colours as possible.

Results. A record of each student prediction was made. Previous studies by Ayres and Way (1999, 2000) found that students often changed their strategy over the last three predictions compared with the first three predictions. Consequently, the number of whites chosen in the first and last three predictions were also reported (see Table 2). A 1-way ANOVA (known v unknown ratios) with repeated measures (first 3 and second 3 predictions) were completed on this data. For the main effect, knowing the sample space made no difference; $F(1, 20) = 1.88$, $p > 0.05$. However, students chose significantly more white balls during first 3 predictions than during the second 3; $F(1, 20)$ 7.12, $p < 0.05$. Consistent with previous research with Ayres and Way (1999, 2000), this particular outcome sequence caused many students to switch from using the "most likely" strategy to strategies based on common misconceptions. Qualitative data (due to space restrictions, this data will be reported more extensively at a later date) also confirmed an increased use of misconceptions over the last three predictions, such as "its yellows turn" and "the colours are forming a pattern". However, nearly half the students (45%) continued with a strategy of choosing the most frequently occurring colour for the last 3 predictions.

Table 2: Mean number of white balls selected in Video Test

	Ratio Known (n = 11)	Ratio Unknown (n = 11)	Combined Group Totals
First 3 Predictions	2.6 (0.5)*	2.3 (0.8)	2.5 (0.7)
Second 3 Predictions	1.9 (0.9)	1.6 (0.9)	1.8 (0.9)
Overall Predictions	4.5 (1.2)	3.9 (0.9)	4.2 (1.1)

*Note: Standard deviations are given in brackets.

To explore the influence of the initial group instructional activities further, prediction means were calculated for each of the instruction groups (see Table 3). Although there were no significant between-group effects (group numbers were very small) on 1-way ANOVAs for these measures, group means did vary considerably. On the crucial measure of the second set of predictions, Groups 4, 5 and 6 chose almost twice as many whites as Groups 1, 2 and 3.

Table 3: Mean number of white balls selected by each instructional group

Instruction Groups	First 3 Predictions	Second 3 Predictions	Total
1	2.7	1.3	4.0
2	2.0	1.3	3.3
3	2.7	1.0	3.7
4	2.5	2.3	4.8
5	2.8	2.0	4.8
6	2.3	2.5	4.8

It was noticeable that there appeared to be a match between the group data for this measure and some of the statistics shown in Table 1. Correlation calculations revealed that the number of whites predicted on this trial was significantly correlated to both the actual number of greens (dominant colour) that occurred during the group activities ($r= 0.45$, $p < 0.05$) and the number of correct predictions made during these activities ($r = 0.55$, $p < 0.01$). Both fairly strong positive values suggest that the differing experiences within groups had an effect which transferred into the video trial.

CONCLUSIONS

The main aim of this study was to test the effectiveness of an instructional period, featuring small group practical activities, on primary aged students' development of probabilistic reasoning. Whereas, students generally showed a good understanding of likelihood, when faced with an unrepresentative set of outcomes (the video trial) many students reverted to strategies based on misconceptions. In a similar fashion to the previous studies of Ayres and Way (1999, 2000), students chose less white balls (the most commonly occurring colour) over the final three predictions compared to the first three, indicating that their use of the "most likely" strategy was reduced. However, it is worth noting that the prediction rates of white in this study were higher than previously found. For example, in the Ayres and Way (2000) study with grade 8 students, overall mean values for this video outcome sequence was 2.0, 1.4 and 3.4, for the first 3, second 3 and total predictions respectively. In the Ayres and Way (1999) study with grade 6 students, the mean values were 1.5, 1.2 and 2.7. Both sets of data were lower than those found in this study (2.5, 1.8 and 4.2). These comparisons suggest that the instructional period in this study, as short as it was, may have helped students developed a better understanding of chance.

Of considerable interest in this study was the group effects which appeared. During the small group activity, students were exposed to different sample spaces, which produced varying sets of outcomes. Some groups were more successful in predicting than other groups, with success rates apparently dependent upon the observed outcomes. Groups which observed a set of outcomes representative of the sample space tended to demonstrate a better understanding of likelihood than groups who observed less likely outcomes. Furthermore, these differences during instruction appeared to influence decision making during the final video trial.

The implications of this study are as follows. Firstly, instruction in this domain seems to have, at least in the short term, a positive effect. Secondly, the types of outcomes observed in these random-based activities, seem to have an influence on future decision making strategies. Consequently, we believe that any instruction of this nature must ensure exposure to different types of outcomes to provide students with the opportunity to develop a clearer understanding of the relationship between sample, randomness and likelihood. Furthermore, because the groups had different experiences, teachers need to be aware of the possibility of collective misconceptions forming within groups.

Finally, it must be acknowledged that this was a small study based on a particular sample of students. It is our aim to investigate these findings further, using a broader sample of students and modified versions of the instructional period.

REFERENCES

Amir, G., Linchevski,L. & Shefet, M. (1999). The probabilistic thinking of 11-12 year old children. In Zaslavsky, O., (Ed.) *Proceedings of the Twenty Third Conference of the International Group for the Psychology of Mathematics Education,* Haifa, Israel, (2) 25-32.

Ayres, P & Way, J (1998). Factors influencing predictions made on randomly generated sequences. In Kanes, C., Goos, M. & Warren, E. (Eds.). *Published Proceedings of the Mathematical Education Research Group of Australasia (MERGA 21),* Gold Coast, Australia, 74 - 81.

Ayres, P & Way, J (1999). Decision-making strategies in probability experiments: the influence of prediction confirmation. In Zaslavsky, O., (Ed.) *Proceedings of the Twenty-third Conference of the International Group for the Psychology of Mathematics Education.* Haifa, Israel, (2) 41-48.

Ayres, P, & Way, J. (2000). Knowing the sample space or not: The effects on decision-making. *Published proceedings of the 24th Psychology of Mathematics Education Conference*, Hiroshima, Japan, (2) 33-40.

Fischbein, E. & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal of Research in Mathematics Education*, *28*: 1, 96-105.

Jones, G., Thornton, C., Langrall, C. & Mogill, T. (1996). Using children's probabilistic thinking to inform instruction. In L, Puig & A, Gutierrez (Eds.), *Proceedings of the Twentieth International Conference for the Psychology of Mathematics Education,* Valencia, Spain, (3) 137-144.

Jones, G., Langrall, C., Thornton, C. & Mogill, T. (1999). Students' probabilistic thinking in instruction. *Journal for Research in Mathematics Education, 30(5),* 487-519.

Shaughnessy, J. M. (1981). Misconceptions of probability: From systematic errors to systematic errors and decisions. In A. Shulte & J. Smart, *1981 Yearbook: Teaching Statistics and Probability*. (pp.90-100.) Reston, Vi.: NCTM.

Truran, J. (1996). Children's misconceptions about the independence of random generators. In L, Puig & A, Gutierrez (Eds.), *Proceedings of the Twentieth International Conference for the Psychology of Mathematics Education,* Valencia, Spain, (4) 331-338.

Watson, J. & Moritz, J. (1998). Longitudinal development of chance measurement. *Mathematics Education Research Journal, 10 (2),* 103-127.