CAN GRADE 3 STUDENTS LEARN ABOUT VARIATION? ®

Jane M. Watson and Ben A. Kelly
University of Tasmania, Australia

This paper will report on outcomes observed in an investigation that involved teaching chance and data with an emphasis on understanding the part that variation plays in processes associated with chance measurement and data collection/analysis. Classes of students in grades 3, 5, 7, and 9 took part in the study but this report will focus on children in grade 3. They were taught a unit of 10 lessons over eight weeks and given pre and post tests in association with the teaching of the unit. Of interest was not only their learning about basic probability and data handling but also their developing understanding of the influence that variation has on outcomes in relation to the observation of pattern. The question of the age at which children can start appreciating the influence of variation creates special interest in this group of students.

INTRODUCTION

Variation is at the heart of all statistical investigation. If there were no variation in data sets, there would be no need for statistics. Although statistical variation is taken for granted by statisticians, school students often have little concept of appropriate variation and many tertiary students also fail to understand the variation behind the formulae they learn to measure it. Traditionally, standard deviation is the common formula used to measure spread; however, due to its complex nature it is often avoided by educators and definitely not included in the primary and middle school mathematics curriculum. The concept of variation, however, is an important element of basic understanding of statistics and "real world" functioning and does not necessarily rely on the understanding of complex formulae to be taught effectively. In fact, by using appropriate materials and activities to introduce variation to younger students, it is thought that perhaps students as young as grade 3 can obtain a realistic appreciation of variation.

Reported here are the results of an investigation carried out with grade 3 students who participated in a unit of work on chance and data designed to emphasize variation. This investigation was part of a larger project designed to provide a model of development of students' understanding of statistical variation across the years of schooling, as well as monitoring change after teaching had occurred. This report will focus on the following aspects of the research conducted with grade 3 students: a description of the sample and the teaching unit in which they participated, a description of the survey used as a pre and post test with the sample, the results of analyzing the pre and post survey with respect to subscales and the survey overall, and a discussion about grade 3 students' facility with the idea of variation.

SAMPLE AND TEACHING UNIT

The sample of grade 3 students in this study consisted of 72 children from three primary schools (n = 32, 19, and 21). There were 30 boys and 42 girls with roughly equal numbers in each school. Two classes of students were taught in each school and each class met with the teacher, a primary-trained mathematics specialist, on 10 occasions over an eight-week period. The purpose of the unit was to cover aspects of the chance and data curriculum appropriate for grade 3, while at the same time emphasizing variation as it occurred during the activities.

The investigation of the contents of a small packet of Smarties™ (small round chocolate sweets of different colours) was chosen for the first session as it was thought that the motivation of a "Smarties™ feast" at the conclusion of the session would maintain interest. The focus was on gathering different types of data so discussion began with the outside of the box. A great deal of information about the contents and the manufacturers was obtained. When the boxes were opened, students working in pairs were asked to "find out" about the contents. All pairs counted and sorted the Smarties™ and developed a column graph of the colours. The discussion centred around the number of Smarties™ in each box, the different colours, and the number of each colour. Variation among the contents of the boxes was noted. For each class a table was constructed to show the numbers of Smarties™ in the boxes. This was converted into a stacked dot plot (referred to as a line plot by Russell & Corwin, 1989, p. 17) and the two representations compared. Students were able to comment on the range and the most common values and to compare their group results with those of the whole class. Another whole-class graph showed the distribution of colours. Several students commented on the fact that there were "holes" (missing colours) in the group graphs but not in the whole-class one. In one class there were far more green Smarties™ than any other colour and it was suggested that green colouring might be cheaper. It became necessary to introduce a scale of "one x for two Smarties™", which was a new idea for some of the students.

The objective for Session 2, about "People in a Family", was to develop ideas about defining the data to be collected, representing the data in different ways and describing the general shape of the data (Russell & Corwin, 1989). After reaching consensus on the definition of a family, three different representations were developed and discussed in terms of suitability for sharing with other groups. A "people graph", where the students stood in columns according to the number of people in their families, was converted to towers of linked blocks and then to a column graph on a large piece of paper. Each student placed a dot representing his/her family on the column graph so that it was possible to locate individual contributions to the class graph. The students felt that this was important. Responses to the question, "From this graph what can we say about the families of the people in this class?" included comments about the range of sizes and the most common size. In one class there was a family of nine. This led to a discussion of values that were a long way from the rest and in the context of discussing variation in family size, the term "outlier" was introduced. At this stage the idea of a mid-point was introduced through the question, "What can we say about half of the class?" Students were also asked to talk about the general shape of the graph. Responses included "pencils", chimneys", and "buildings". The students were looking at the individual columns and hence a line was drawn around the graph to stress the overall appearance. Comments changed to "a mountain", "a roller coaster", "a giant rock", and "a rocket". The idea of sharing the information with others led to the recognition of the need to label the graphs with the class name and the date.

Sessions 3 and 4 were on chance. The first dealt with equiprobable events using a spinner and a single die, whereas the second considered events that were not equiprobable using two dice and summing outcomes (Edwards & Hensein, 2000; Watson, 1994). The students were quick to grasp the idea of quantifying chance in terms of a single event with the die or the spinner, although some still appeared to have a lingering suspicion that six "really" was the hardest number to get. Students compared the class graphs generated with one die ("a box", "a table", "a flattish hill") with those generated with two dice ("a big hill", "like the families"). The idea of the "law of large numbers" was introduced and some students indicated their understanding of chance with comments such as "Things aren't perfect, you know!" and "If you get them all the same [with one die] you probably cheated!" In one class, for all of the following sessions, the teacher arrived to find on the blackboard, "What is the chance of Mrs Jeffery wearing stripes? …having a black handbag? etc.", with the students' suggestions written underneath. In another group there was a discussion of the chance of the teacher's car having a flat tyre on the way home. As soon as she arrived for the next session some days later, she was met with, "Did you have a flat tyre?"

The idea of sampling to find out about a population, the focus of Sessions 5 and 6, stimulated a lively discussion. The students in each group were able to give many examples of situations where sampling occurs (e.g., cosmetic counters, tasting in cooking, supermarket displays, wool classing, etc.) and were able to suggest why a sample is tested rather than the whole. Each class came up with similar definitions of sampling as "testing a little bit of something that tells you what the whole thing is like when it is not possible to test the whole thing." The selection of a sample representative of the population was an issue for the students. As each member of the population (class) had to have an equal chance of being chosen in the sample, playing cards and a die were used for selection purposes. The children became very interested in "fairness" and when the same students were selected in different samples it was considered not fair because they had already had a turn. The idea of predicting from a sample to a population was the most difficult of those dealt with throughout the project. Most students had difficulty with proportional reasoning – not surprising for grade 3. The ideas of sample size and replacement after each trial also proved difficult. A feature of the unit was the sampling of cubes of two colours from opaque bags, recording results, and comparing expectations with what actually occurred—variation was a part of the discussion (Corwin & Friel, 1990; Watson, 1994).

Sessions 7 and 8 were based on students' measurement of how long they could stand on each foot with their eyes closed (Rubin & Mokros, 1990). The purpose was to introduce an experimental method that would generate data and lead to the use of stacked dot plots to compare sets. Each student was given a green dot to represent the right foot and an orange dot to represent the left foot. After agreeing to the rules, students worked purposefully in groups with very little disruption to collect the data. The dots were entered onto stacked dot plots which were then analysed in terms of clusters, holes, unusual scores, the most common score and the range. The students were able to compare these aspects of the two line plots and reach conclusions about the differences between the two data sets.

The final two sessions (9 and 10) were planned to give students the opportunity to use the ideas that were introduced during the unit to plan their own investigations. Each student was given a pencil, measuring equipment, and a recording sheet. Each class decided on a question to be answered by collecting data in relation to how far a pencil could be blown on a flat surface. Some of the questions were: "In our class is there a difference in pencil-blowing power between: the people who play sport and those who don't? …girls and boys? …cross-country runners and others?" Setting the rules for data collection was time consuming but felt to be very important by the students. Data were collected efficiently and collated on a whole class basis. Analysis of the data required a great deal of teacher direction and in all cases the results were inconclusive in terms of group differences. The methods chosen were perhaps too sophisticated for many grade 3 students with limited experience but many others appeared to appreciate the process involved.

SURVEY INSTRUMENT AND ANALYSIS

The same pre and post survey was given to the students. It was a subset of items administered to older students due to the complexity of some concepts and the required reading levels. Besides including some items on basic chance and data skills, such as table and graph reading, three themes relating to variation were identified as being important for an understanding of statistics: variation in chance, variation in data and graphs, and variation in sampling situations.

Traditional chance and data concepts were addressed using items adapted from earlier studies. Basic chance was covered in questions about the outcome of rolling a six-sided die, about the likelihood of choosing a marble of a certain colour from one of two boxes of marbles (Watson, Collis, & Moritz, 1997), and an item using a 50-50 spinner (Torok, 2000). Four questions involved simple table reading exercises based on a school sports day scenario, adapted from Watson (1998), whereas two others were simple pictograph reading exercises based on how children travel to school, adapted from Watson and Pereira-Mendoza (1996).

The 50-50 spinner context of Torok (2000) was further developed to investigate variation in chance by asking for a prediction of the outcomes (i) for 10 spins, (ii) for a repeated trial of 10 spins, (iii) that would be surprising (e.g., 9 out of 10 spins), and (iv) for six sets of 10 spins. Another question concerned the predicted outcomes of 60 tosses of a die and students were asked to fill in a table. Variation in data and graphs was assessed through questions about the pictograph, which involved possible variation from day to day and inference or prediction based on the information displayed (4 items). Variation in sampling was assessed through an extension of the table-reading question that asked for fair ways of selecting children to lead a parade and through a question that asked for a definition and an example of a sample. As well, five questions were based on the "raffle scenario" of Jacobs (1999), which related to a class planning a survey for its school before selling raffle tickets. One question asked for the student's suggestion and four others asked for evaluation of specific methods, some of which were biased.

The posttest was administered approximately seven weeks after the completion of the teaching unit at all schools. For the purpose of this paper, five variables are defined based on the survey. One is the sum of scores on all items and four are based on the subscales relating to basic chance and data, variation in chance, variation in data, and variation in sampling. The scoring for each item ranged from 0-1 to 0-5 depending on the potential sophistication or complexity of the response. The scoring rubric was the same as that used for all other grades hence it was possible that no grade 3 obtained the highest score for a particular item. Of interest in this study is the change, potentially the improvement in scores from the pre to the post survey. Paired t-tests were performed for the five variables (n = 72).

RESULTS

Pre and post survey means and standard deviations are given in Table 1. As can be seen, statistically significant improvement occurred on the four subscales and on the overall survey. Several comments are appropriate on particular items that contributed to these scales.

Table 1
Paired T Tests

Scale	Pre Mean (SD)	Post Mean (SD)	*t*	*P*
Basic Chance and Data (9 items)	9.7 (2.8)	10.8 (2.4)	-3.7	<.001
Variation in Chance (5 items)	4.1 (2.8)	5.5 (2.6)	-5.3	<.001
Variation in Data/Graphs (4 items)	3.9 (1.8)	4.4 (1.9)	-2.6	<.01
Variation in Sampling (7 items)	5.3 (3.6)	6.5 (4.8)	-2.7	<.01
Total (25 items)	23.0 (8.6)	27.2 (9.8)	-6.4	<.001

On only two items did students perform on average better on the pretest than the posttest. In each case the difference was not significant. For one item on variation in data students were asked what a gap in the travel graph in a row associated with a train meant. Although there was no "correct" answer to the question, the use of more statistically appropriate reasoning such as "You could go by train but no one did" was uncommon on the pretest and totally absent on the posttest. Basic interpretations like "They don't catch the train" were more common on the posttest than the pretest to the detriment of more sophisticated reasoning. The other item, on sampling, asked students which of three suggested survey methods was the best. It is most likely that the drop in average score was because many students did not answer this item, the last, on the posttest. This may have been due to students spending more time thinking about and explaining responses to earlier items because they had more to offer.

The items where students improved the most with respect to basic understanding of chance and data included the spinner item with a probability of 1/2 landing on black, naming how many people walked to school in a pictograph (which all students got correct in the posttest), recognizing the most popular sport from a table, and calculating the total number of children at a sports day from entries in a table.

Significant improvement occurred for all items in the subscale on variation in chance reflecting increased appreciation for the distribution of outcomes from tossing a die 60 times and expectations when spinning a 50-50 spinner, particularly 10 times on six different occasions. A student on the die question, for example, went from an inappropriate distribution and response in the pretest ("2, 8, 45, 0, 3, 2 – Because they sometimes come up when I roll the dice") to a strict probabilistic understanding and outcome on the posttest ("10, 10, 10, 10, 10, 10 – Because they have an even chance"). Another student improved from a strict probabilistic outcome (10, 10, 10, 10, 10, 10 – Because each number has an even chance of coming up"), to an understanding of proportion and appropriate variation on the posttest ("12, 8, 10, 9, 11, 10 – Because they are all around 1/6 of 60 which is 10 and there are 6 numbers on the die"). In the spinner question involving the repetition of 10 spins 6 times, one student went from a wide prediction of variability in the pretest ("3, 9, 1, 5, 10, 0") to an appropriate prediction of variability in the posttest ("4, 6, 7, 3, 2, 4"). It is of interest to note the elimination of the extreme values in the posttest prediction.

The greatest change for items involving variation in data or graphs occurred for an item based on the pictograph that asked for a prediction of whether a new child in the class who came by car was a boy or a girl and how the decision was made. In this case, students on the posttest were more likely to respond based on the information in the pictograph and acknowledge potential variation in the answer. Changes in student responses from no attempt or no reasoning behind the prediction in the pretest, to a prediction based on patterns ("Boy, because it is a pattern") or inverse majority [balancing] ("Boy, because there is one boy") on the posttest were common. Some students managed to improve further by providing an inference from the data in the posttest ("Girl, because more girls take the car"). Another student responded using inverse majority [balancing] in the pretest ("Boy, there are less boys that come by car"), but in the posttest made an inference and appreciated the variation and uncertainty ("I don't know, probably a girl. There's more girls that go by car"). This response was in the highest category defined.

For the subscale on variation in sampling, significant changes occurred in the item that suggested a random method for surveying children in a school with students accepting this idea more readily on the posttest than the pretest. As well, students improved significantly in defining a sample and in describing fair methods of selecting children to lead a closing parade at a sports day. Typical responses for the random method of surveying improved from not understanding the question or context in the pretest ("Not sure, because I don't think it is very good putting all the names in a hat"), to a non-central yet appropriate appraisal of the method in the posttest ("Good, because it is fair"). Very few students in either test, however, responded at a level that was indicative of a sophisticated understanding of the benefits of a randomly drawn sample. When asked to give a definition of sample, most students improved from not attempting the question, to showing an understanding of the part-whole relationship in the posttest ("It means a bit of something"). One student went from giving a single idea and somewhat unsure definition ("Taking something to get the answer to your question") to a sophisticated definition in the posttest ("To take something out of something bigger to test"). In regards to the question asking students to describe fair methods of selection, most went from using representative methods in the pretest ("2 girls and 2 boys"), to using random methods in the posttest ("Draw them out of a hat"). A couple of students managed to take this further by combining representative and random methods in the posttest ("Pick 2 boys out of a boy hat and 2 girls out of a girl hat"). This was the highest category of response.

DISCUSSION

Although acknowledging that the grade 3 students in this study did not reach the levels of achievement in the posttest that were attained by older students and that many students continued to have difficulties, the improvement observed indicates that progress can be achieved in these children's basic understanding, as well as their understanding of variation. The improvement in the subscale on basic understanding of chance and data is undoubtedly due to the exposure to these ideas during the 10 teaching sessions. Discussion of events and their probabilities, and reading from graphs and tables featured incidentally throughout the unit.

As can be seen in Table 1 the greatest improvement with respect to variation occurred for variation related to chance. The activities in the unit involved both dice and spinners and although they were not closely tied to the questions in the survey, students would at least be familiar with the context and perhaps more willing to have an attempt at questions. Activities in the unit associated with sampling, data recording and summarizing did not use the same contexts as the questions in the surveys and hence it is not surprising that the transfer and improvement were not as great.

The students were quite enthusiastic about the activities undertaken and in some classes the usual classroom teacher reinforced the ideas suggested by the project teacher in between visits. This was shown, as noted above, by the questions prepared for her arrival. The chance activities were particularly popular as the children enjoyed predicting outcomes and then checking them. Their appreciation of uncertainty and variation was shown in the comments quoted earlier. One instance was observed during classroom activities where a pair of students cheated to make their outcomes fit with the theory exactly. This suggests teachers must motivate as much interest in variation from the expected pattern in chance outcomes, as in the expected pattern itself. Judging by the class discussion this was the case for most children.

Although the children had difficulty with the proportional reasoning necessary to make inferences from samples to populations, they had good intuitions about samples from their out-of-school experiences. These were built upon during class discussions and the improvement was evident on the posttest. It hence appears important to begin talking about sampling at this level, which should assist understanding later when other mathematical skills have developed further. The issue of fairness in sample selection is one also noted by Jacobs (1999) in her study of middle school children. Jacobs suggests that children often devise or evaluate methods of selection based on a fairness rationale that relies heavily on emotive and personal beliefs rather than on traditional statistical ideas of what constitutes "fair" (e.g., how others will feel when they are/are not selected versus giving everyone an equal chance of being selected). This fairness rationale can quite often lead to inaccurate evaluations, especially when applied to truly randomized selection procedures.

At grade 3, students were especially interested in collecting data about themselves. As noted by the project teacher, they were keen to record their own dots on the class graphs and hence see where they were compared to others in the class. This ownership of data was highly motivating. The activities where students collected data on how long they could stand on each foot with their eyes closed or blow a pencil across a flat surface encouraged aspects of planning that motivated the students. These sophisticated ideas, however, were not easily questioned on the survey. The motivation associated with comparing two sets of data (right and left foot, or boys and girls) led to much excitement when discussion took place, reinforcing calls by Watson and Moritz (1999) and the National Council of Teachers of Mathematics (2000) for activities of this type to be included in the primary school mathematics curriculum.

The question in the title of this paper was, "Can grade 3 students learn about variation?" This study has shown that the answer is "yes". Although grade 3 students did not reach the highest level of achievement possible, the pre-post test analysis indicated significant improvement in understanding. This improvement was not only statistically significant but also educationally significant as is illustrated in the responses given. It is to be hoped that this study will encourage curriculum writers to include variation in the chance and data expectations for grade 3 students. This will help lay a firm foundation for the later study of statistics.

ACKNOWLEDGEMENTS

This research was funded by an Australian Research Council grant (No.A00000716). The authors wish to thank Patricia Jeffery, the project's energetic and innovative classroom teacher.

REFERENCES

Corwin, R.B., & Friel, S.M. (1990). *Used numbers. Statistics: Prediction and sampling.* Palo Alto, CA: Dale Seymour Publications.

Edwards, T.G., & Hensein, S.M. (2000). Using probability experiments to foster discourse. *Teaching Children Mathematics, 6*, 524-529.

Jacobs, V.R. (1999). How do students think about statistical sampling before instruction? *Mathematics in the Middle School, 5*, 240-263.

National Council of Teachers of Mathematics (2000). *Principles and Standards for School Mathematics*. Reston, VA: Author.

Rubin, A., & Mokros, J. (1990). *Data: Kids, Cats, and Ads: Statistics*. Menlo Park, CA: Dale Seymour Publications.

Russell. S.J., & Corwin, R. (1989). *Used Numbers: The shape of data*. Palo Alto, CA: Dale Seymour Publications.

Torok, R. (2000). Putting the variation into chance and data. *Australian Mathematics Teacher, 56*, 25-31.

Watson, J. (1994). *Maths Works: Teaching and learning chance and data*. (Assisted by W. Ransley) Adelaide, SA: Australian Association of Mathematics Teachers.

Watson, J.M. (1998). Numeracy benchmarks for years 3 and 5: What about chance and data? In C. Kanes, M. Goos, and E. Warren (Eds.), *Teaching Mathematics in New Times* (Vol. 2, pp. 669-676). Brisbane, Q: Mathematics Education Research Group of Australasia.

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. (1999). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics, 37*, 145-168.

Watson, J., & Pereira-Mendoza, L. (1996). Reading and predicting from bar graphs. *Australian Journal of Language and Literacy, 19*, 244-258.