GRADE 5 STUDENTS' APPRECIATION OF VARIATION

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

This report focuses on one aspect of a larger study of school students' understanding of statistical variation. Although the study included students in grades 3, 5, 7, and 9, this paper will focus on grade 5 students only. Students experienced a unit of 10 lessons on the chance and data part of the mathematics curriculum conducted over an eight-week period. Lessons included a particular emphasis on variation and its role in statistical understanding. Pre- and post-tests were administered and improvements were found in overall performance and for variables reflecting appreciation of variation in chance, variation in data, and variation in sampling. Some comparisons are made with grade 3 students' performance.

INTRODUCTION

Variation is an important element in the basic understanding of statistics (Moore, 1990), yet many middle school students have little understanding of what constitutes appropriate variation. Partly because of the complex formula traditionally used to measure variation, educators and curriculum developers tend to avoid it at this level. As a result, students usually do not encounter the concept of variation or its applications until senior level mathematics in school or university statistics courses. Until this time there is a tendency to focus on centers with little regard for spread or variation within data sets. It is possible, however, to help primary school students reach a broader appreciation of variation without the computation of complex formulas. In a related investigation by the same researchers (Watson & Kelly, 2001) it was shown that students as young as grade 3 could develop a foundation for the understanding of some aspects of statistical variation. Although this understanding was limited, it was hypothesized that slightly older students, with the same instruction and materials, could gain a firmer grasp of the concepts involved.

The data used in this study were part of a larger project designed to model development of, and monitor change in, students' understanding of statistical variation in grades 3, 5, 7, and 9. This report will focus on the change in understanding following instruction for grade 5 students, comparing improvements with grade 3 students in the same study (Watson & Kelly, 2001). Similar patterns of improvement and an understanding at a slightly higher level were expected.

SAMPLE AND TEACHING UNIT

The sample of grade 5 students in the present study consisted of 82 children from three primary schools in the Australian state of Tasmania (n = 35, 31, and 16). Students were taught 10 lessons covering aspects of the chance and data curriculum with a special emphasis on variation. All lessons were taught by the same experienced teacher, provided by the research project.

The first lesson was an investigation of a small packet of coloured chocolates called Smarties™. Students were asked to investigate the contents of the packets in pairs or groups and develop column graphs to show the number of different colours in each box. Whole-class stacked dot plots were created from the individual graphs and these were compared against the data collected in individual pairs. Lesson 2 was about "People in a Family" (Russell & Corwin, 1989), and students recorded the number of people in their families. Once again, different types of graphs were created; people graphs, linked blocks, and paper-and-pencil graphs were used to display the information. The types of information that can be obtained from such representations, the general shape of the graphs, and the existence of outliers were discussed during this session. The third and fourth lessons dealt with equiprobable events using spinners and dice, and non-equiprobable events using two dice. The activities allowed students to compare class graphs and discuss the shapes of graphs generated from one die with graphs generated from two dice (Edwards & Hensein, 2000). Variation was a strong theme during these sessions, linked with discussions of the theoretically expected outcomes. Lessons 5 and 6 were related to sampling and the first was introduced with a discussion of what constitutes a sample. Activities included selecting randomly drawn samples from the classroom and predicting to the wider classroom population. The issue of fairness and the idea of predicting from a sample were the most difficult concepts for students to grasp.

The final four lessons were designed to introduce students to experimental methods. In lessons 7 and 8 students investigated how long they could stand on each foot with their eyes closed (Rubin & Mokros, 1990). The two data sets (for left and right feet) were then represented on a whole-class-generated stacked dot plot for discussion and analysis. Students focused on clumps, outliers, common scores, range, and the differences between the two sets of data. The final lessons followed on from this and gave an opportunity for students to plan and conduct their own experiments. All students were told they would be given a pencil, measuring equipment, and a recording sheet, and would be asked how far the pencil could be blown across a flat surface. Each class then had to decide on a question and develop a hypothesis. Most classes focused on possible differences between boys and girls, sports players and non-sports players, and so on. Even though the analysis of the data required a great deal of teacher direction, the students conducted their experiments efficiently. Once again, discussions focusing on differences, common scores, outliers, and the shapes of the graphs reinforced the notion of variability in the data sets.

SURVEY INSTRUMENTS AND ANALYSIS

The pre- and post-tests consisted of the same 29 items, covering four aspects of the curriculum as emphasised in the teaching unit: Basic Chance and Data, Chance Variation, Data Variation, and Sampling Variation. The items on Basic Chance and Data were adapted from the studies of Watson, Collis, and Moritz (1997) involving dice and coloured marbles, of Torok (2000) involving spinners, of Watson (1998) involving basic table reading, and of Watson and Pereira-Mendoza (1996) involving reading pictographs. For Chance Variation there were items involving spinners over many trials developed by Torok (2000), as well as a question concerning the prediction of outcomes for 60 tosses of a die. Data Variation was assessed through pictograph items adapted from Watson and Pereira-Mendoza (1996), as well as three items asking for comparison of two differently scaled stacked dot plots representing the same data on how long students in a class had lived in town (Konold & Higgins, in press; see Figure 1). The items asked students what they could tell from each of the plots and then asked for a decision as to which stacked dot plot told the story better. Sampling Variation items included giving a definition of "sample"; suggesting a method of sampling, and evaluating the bias in four sampling methods, for the "raffle scenario" of Jacobs (1999); and suggesting a fair way to select students to lead a parade in the context of the table reading items of Watson (1998). For a more detailed description of the items and their contribution to the four variables (Basic Chance and Data, Chance Variation, Data Variation, and Sampling Variation), refer to Watson and Kelly (2001).

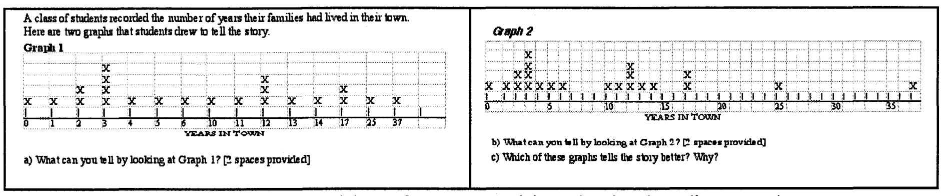

Figure 1. Stacked dot plot problem from Konold and Higgins (in press).

The post-test was administered approximately seven weeks after the completion of the teaching units. Five variables were defined, representing the four aspects of the curriculum covered in the subscales described above, and the total score for the test. The scoring for each item ranged from 0-1 to 0-5 depending on the potential sophistication or complexity of the response. Of interest in this study is the change, potentially the improvement, in scores from the pre- to the post-test and the improvement of performance as compared to the grade 3 students. Paired t-tests were performed for the five variables ($n = 82$). The tests for grade 3 were the same as for grade 5 except that they did not include the three items on stacked dot plots (Konold & Higgins, in press) or one of the biased sampling items (Jacobs, 1999).

RESULTS

Pre- and post-test means and standard deviations for the five variables defined for the grade 5 students are given in Table 1. As can be seen, there is a statistically significant improvement for the three themes of variation and on the overall total. The paired t-test for pre- and post-tests for the Basic Chance and Data variable was not significant. The maximum possible score on this variable was 16 and as can be seen, the grade 5 mean pre-test understanding of Basic Chance and Data was relatively high. The non-significant change may have been due to a ceiling effect.

Table 1. **Paired t tests for the grade 5 survey**

Variable (number of items)	Pre Mean (SD)	Post Mean (SD)	*t*	*P*
Basic Chance and Data (9 items)	12.1 (1.9)	12.4 (2.1)	-1.3	NS
Chance Variation (5 items)	5.6 (2.6)	6.7 (2.1)	-4.3	<.001
Data Variation (7 items)	8.4 (2.9)	9.3 (3.0)	-3.1	<.01
Sampling Variation (8 items)	8.7 (4.0)	11.2 (4.8)	-5.2	<.001
Total (29 items)	34.8 (8.5)	39.6 (9.6)	-6.2	<.001

For the Chance Variation variable students improved on 3 out of the 4 spinner items. Two items showing significant differences required a prediction of how many times the spinner would land on the shaded part out of 10 spins and asked how many spins landing on the shaded part out of 10 would be surprising. In response to the first, a student in the pre-test wrote, "6 times, [because] you spin and hope it will come up", whereas in the post-test she used a more theoretical prediction, "5 times, because they have an equal chance." A few grade 5 students on the post-test were able to expand on this and acknowledge the role of variation. An optimal response was "4, 5, or 6, because it would land on the shaded part approximately half the time." On the spinner item that asked for a prediction of a shaded outcome for a 50-50 spinner spun 10 times on six different occasions, one initial response indicated a strict probabilistic view ("5,5,5,5,5,5"), whereas the corresponding response on the post-test acknowledged the potential for variation over the six trials ("5,3,6,4,7,5").

For the Data Variation variable, all three items comparing the two differently scaled stacked dot plots in Figure 1 revealed a significant improvement from pre- to post-test. One student, for example, in the pre-test responded by reading the data on the first stacked dot plot incorrectly; she wrote, "Most people have lived here one year, and the person who has been here the most has been here for four years." This student mistakenly read the X's on the stacked dot plot to be the years (instead of people) and the numbers on the horizontal axes to be individual people (instead of years) even though it was labelled. The same student responded in the post-test by stating "3 years has the most; and, 0, 1, 4, 5, 6, 10, ... all have one X on them." Although this student did not put the data into the full context, she read and interpreted the horizontal axis correctly, an improvement over the earlier response.

The pictograph items, part of the Data Variation variable, showed no improvement on the post-test. Of interest was the item about the gender of a new student if the new student arrived by car, which required students to make an inference based on the data presented in a pictograph. Statistical reasoning based on the observed majority (girls), accompanied with an element of uncertainty, for example, "Girl, I don't know [why], I'm having a guess ... because there is more chance of it being a girl," was shown by only 1.2% of grade 5 students on the post-test. None answered like this on the pre-test and the majority of students both times responded by focussing on patterns in the data, with no emphasis on uncertainty or observed majorities. The possible reasons for this are discussed later.

For Sampling Variation, there was a significant improvement in performance on five of the eight items: judging two of the survey methods presented in the "raffle scenario", choosing the best survey method out of the four presented, defining a sample, and suggesting fair selection methods for students to lead a parade. Overall, students demonstrated a better post-test understanding of what a sample is and applied this knowledge to examine critically the survey methods presented and suggest methods of selection on their own.

Table 2 shows the means and standard deviations for grades 3 and 5 for the variables defined from the subset of the 25 common items completed by both grades. As can be seen, the grade 3 students improved significantly on the Basic Chance and Data variable, whereas the grade 5 students improved very little. The grade 5 pre-test mean, however, was significantly higher than the grade 3 post-test mean ($p<0.001$) and as noted there may have been a ceiling effect for grade 5. Similarly, for the common items in the Data Variation variable, grade 3 students improved significantly whereas the grade 5 students showed no improvement. Again the grade 5 pre-test mean was higher than the grade 3 post-test mean ($p<0.05$), but this difference was not considered meaningful given the small number of common items.

Table 2. **Comparison of paired t tests of survey results for the grades 3 and 5**

	Grade 3*				Grade 5			
Variable (number of common items)	Pre Mean (SD)	Post Mean (SD)	*t*	*P*	Pre Mean (SD)	Post Mean (SD)	*t*	*P*
Basic Chance / Data (9)	9.7 (2.8)	10.8 (2.4)	-3.7	<.001	12.1 (1.9)	12.4 (2.1)	-1.3	NS
Chance Variation (5)	4.1 (2.8)	5.5 (2.6)	-5.3	<.001	5.6 (2.6)	6.7 (2.1)	-4.3	<.001
Data Variation (4)	3.9 (1.8)	4.4 (1.9)	-2.6	<.01	5.1 (1.5)	5.0 (1.6)	0.9	NS
Sampling Variation (7)	5.3 (3.6)	6.5 (4.8)	-2.7	<.01	7.8 (3.6)	10.1 (4.1)	-5.6	<.001
Total (25)	23.0 (8.6)	27.2 (9.8)	-6.4	<.001	30.6 (6.9)	34.1 (7.8)	-5.6	<.001

*Data from Watson & Kelly (2001)

The Chance Variation variable showed a significant improvement for both grades 3 and 5. The grade 5s performed at a higher level than the grade 3s in both the pre- and post-tests but the post-test mean for the grade 3 students was equivalent to the pre-test mean of the grade 5 students. The analysis of the Sampling Variation common items showed a significant improvement from the pre-test to the post-test for both grades 3 and 5, with greater improvement for the grade 5 students.

DISCUSSION

Two aspects of the outcomes of the research with grade 5 students will be covered in this section: the overall improvement in their performance following instruction and their performance relative to that of grade 3 students. Finally some suggestions will be made for future research.

The improvements observed for the variables reflecting the three themes of variation are almost certainly a result of exposure to experiences reflecting variation in the chance and data lessons taught. The Data Variation variable showed a significant improvement only when the three items from Konold and Higgins (in press) were part of the variable. The improvement on all three items provides evidence of the emphasis on collecting and displaying data in stacked dot plots within the teaching unit presented. Lessons 7 to 10 dealt explicitly with gathering data and generating stacked dot plots from the data. In relation to this, Konold and Higgins state that it is important for students not only to be able to read and interpret data from graphs but also to be able to refer the data back to the context in which they are set. An integrated and optimal response to this item would thus include the context of the question (i.e., how long students have lived in town). In response to the second stacked dot plot, one student, after incorrectly reading the plot in the pre-test, responded in the post-test by saying, "A lot of people have lived in the town between 0 and 5 years." This post-test response not only summarizes the data correctly, but also refers the data back to the context within which the task is embedded, as is recommended by Konold and Higgins.

The greatest improvement occurred in the variable measuring Sampling Variation. Lessons 5 and 6 of the teaching unit focused specifically on sampling. In the first session the students demonstrated pre-instructional knowledge of sample through giving examples encountered in everyday life. Anecdotal evidence from videos of the lessons indicates most students had a good intuitive basis on which to build an understanding of sample (e.g., "a lady giving out food in a supermarket"; "a doctor takes a sample of your blood"). Jacobs (1999) in her study of students in grades 4 and 5 found a similar occurrence, stating that her students had a sound knowledge of sampling before instruction began and were using examples containing elements of statistical samples when asked for a definition. The objective of the formal work with students in this study was to improve on the developing understanding of sample by putting it into different contexts and highlighting its potential for description and prediction. The results in Table 1 highlight the significant improvement in the understanding of sampling after the teaching unit.

In comparison to the grade 3 students, the grade 5 students performed at a higher level on the Basic Chance and Data variable and on the common items in the Data Variation variable, even though they did not improve significantly after instruction (Table 2). Anecdotal evidence from preliminary analysis of grade 7 and 9 data shows that the pre and post means for these grades were equivalent to or slightly lower than those reported for the grade 5s, supporting the idea of a ceiling effect for these items. The lack of improvement on the common items for Data Variation for grade 5, however, may be related to mathematical experiences in the classroom outside of this study. On the tests the items used were based on interpreting a pictograph, whereas during teaching stacked dot plots were used more frequently. Patterns of various types (e.g., girl, girl, boy, girl, girl, boy) were present in the pictograph data and recognising such patterns as significant is part of most primary school mathematics programs. Responses to the pictograph items often reflected pattern recognition and whereas for grade 5s this was consistent across the pre- and post-tests, for grade 3 it occurred more in the post-test indicating an "improvement" from not being able to interpret the graph at all. The lack of specific discussion of pattern and pictographs during the lessons meant that students were not specifically encouraged to move beyond this level in either grade. For the grade 3 students—who were just starting graph reading and interpretation—any experience during the lessons was helpful in improving their performance. Grade 5 students had previous experience and began at a slightly higher level on the pictograph interpretation tasks. The lack of discussion of pictographs meant their performance did not change on these items (Table 2) but they reacted well to stacked dot plots, and hence improved on the Data Variation variable with more items (Table 1).

In relation to the Sampling Variation subscale, grade 5 students made a larger absolute gain in this area than the grade 3 students. Students in grade 3 had good out-of-school experiences (e.g., were able to give reasonable examples of a sample) but struggled with making inferences from samples to populations (Watson & Kelly, 2001). This made it particularly difficult for the grade 3 students to choose which of the survey methods was the best in the "raffle scenario". Perhaps the greater initial understanding demonstrated by the grade 5 students led to adopting the concepts and applying them to an unfamiliar context with greater ease and to an improvement in the ability to judge the more appropriate survey methods.

Overall, the difference in average levels of performance favouring grade 5 students over grade 3s would be expected due to their greater time in school and experience in everyday life, as well as their potentially greater levels of cognition and better ability to comprehend the objectives of the lessons. Although the teacher made every effort to conduct each lesson on the same topic in a similar fashion, student input sometimes influenced the direction of the lessons. In grade 3 this sometimes resulted in the reinforcement of ideas already presented, and for the grade 5 students it sometimes led to an extension of the original ideas. All in all, what was encouraging was that within the expectations for the grade levels, both groups improved in their understanding of aspects of variation with chance and data.

Three directions for future research are suggested by the outcomes of this study. First is an analysis of how much transfer is to be expected between the content of the lessons taught and the content of the pre- and post-tests. Obviously, using identical content in the tests and lessons is likely to lead to statistically significant but educationally less useful results. On the other hand if the differences are too great, as perhaps was the case with pictographs and stacked dot plots, then some alterations need to be made after clarifying objectives. Second, research into the performance of other groups after instruction would increase the depth of understanding of how elementary and middle school children understand variation. Are there differences in the impact of instruction between private and public schools in pre and post understanding? Do samples from other countries that perceive chance and data to be important behave similarly to this Australian sample? Is there a universal developmental pattern emerging for a concrete understanding of variation in chance and data concepts? Third, there is the issue of using classroom teachers as the conveyors of the lessons in future research. In this study, classroom teachers were reluctant to organise and teach lessons themselves. Since teachers' own competencies and understanding of the topic are important determinants of the successful teaching of that topic, professional development is strongly advised for those taking part. Teacher attitude, enthusiasm, and topic knowledge, may affect pre-post results.

Acknowledgments: Research funded by an Australian Research Council grant (No. A00000716). The authors thank Patricia Jeffery, the project's classroom teacher.

REFERENCES

Edwards, T.G., & Hensein, S.M. (2000). Using probability experiments to foster discourse. Teaching Children Mathematics, 6(8), 524-529.

Jacobs, V.R. (1999). How do students think about statistical sampling before instruction? Mathematics in the Middle School, 5(4), 240-263.

Konold, C., & Higgins, T.L. (in press). Working with data: Highlights related to research. In D. Schifter, V. Bastable, & S.J. Russell (Eds.), Developing mathematical ideas: Collecting, representing, and analyzing data. Parsippany, NJ: Dale Seymour Publications.

Moore, D.S. (1990). Uncertainty. In L.A. Steen (Ed.), On the shoulders of giants: New approaches to numeracy (pp. 95-137). Washington, DC: National Academy Press.

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

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

Torok, R. (2000). Putting the variation into chance and data. Australian Mathematics Teacher, 56(2), 25-31.

Watson, J.M. (1998). Numeracy benchmarks for years 3 and 5: What about chance and data? In C. Kanes, M. Goos, & E. Warren (Eds.), Teaching mathematics in new times (Vol. 2, pp. 669-676). Brisbane, QLD: 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(1), 60-82.

Watson, J.M., & Kelly, B.A. (2001). Can grade 3 students learn about variation? Paper submitted to the Sixth International Conference on Teaching Statistics, Durban, South Africa.

Watson, J., & Pereira-Mendoza, L. (1996). Reading and predicting from bar graphs. Australian Journal of Language and Literacy, 19(3), 244-258.