ISSUES FOR STATISTICAL LITERACY IN THE MIDDLE SCHOOL

Jane M. Watson
University of Tasmania, Australia
Jane.Watson@utas.edu.au

Focusing on the word "literacy" in the phrase "statistical literacy," the present study explored what happened to the non-numerically based aspects of statistical literacy when students in Grades 7 and 9 were exposed to a unit of work in chance and data that emphasized variation. To test the suggestion of transfer of thinking skills to the literacy side of statistical literacy, 20 items from a larger survey were selected, upon which changes in literacy skills could be measured. Ninety students in each of Grade 7 and Grade 9 were asked the questions in a longer survey before and six weeks after taking part in a unit on chance and data devised by their usual classroom mathematics teacher as part of their schools' mathematics programs.

INTRODUCTION

The use of the word "literacy" in the phrase "statistical literacy" implies a reliance on reading and writing skills associated with critical thinking and communication. Combining "literacy" with "statistical" implies that some statistical skills are also involved and at the school level this conjures up thoughts of calculations such as those associated with means, standard deviations, medians, and various simple and less simple probabilities. Virtually no research, however, has focused only on the literacy aspect without including numerical skills. Anecdotal evidence suggests that students consider questions without numbers in them as "not mathematics" and hence not as important as questions requiring computations. A complete statistical investigation, however requires considerable literacy skill in describing sampling methods and analysis procedures, as well as providing a case for the inference drawn at the conclusion. The critical thinking aspect of statistical literacy, acknowledged by many (e.g., Gal, 2002; Watson, 1997), also requires critical literacy skills in being able to argue or document a case against another's findings. The present study seeks to address the issue of critical literacy within the context of a classroom setting where students are exposed to a unit of work focusing on variation and its place in the chance and data curriculum. The middle years (Grades 7 and 9) are chosen because this is the time when literacy skills generally should be being consolidated and students should have been exposed to social contexts where they can be expected to develop a questioning attitude. The special focus of the present study is the statistical literacy skills of students before the unit of work and the change that occurs and is measured six weeks after its completion.

BACKGROUND

Statistics and Probability have only been officially part of mathematics curriculum documents in English-speaking countries since around 1990 (e.g., National Council of Teachers of Mathematics, 1989). Statistical literacy as such, however, is not mentioned in curriculum documents, with New Zealand perhaps coming the closest in titling one of its three statistics strands "Interpreting Statistical Reports" (Ministry of Education, 1992). The importance of statistical literacy when students leave school, however, was noted by Wallman (1993) in her Presidential Address to the American Statistical Association in the early 1990s. Throughout the decade Iddo Gal also addressed the statistical literacy needs of adults, culminating in a significant review in which he explicitly stated his expectation for adults in the following terms; they should be able

(a) to interpret and critically evaluate statistical information, data-related arguments, or stochastic phenomena, which they may encounter in diverse contexts, and when relevant,
(b) to discuss or communicate their reactions to such statistical information, such as their understanding of the meaning of the information, their opinions about the implications of this information, or their concerns regarding the acceptability of given conclusions. (Gal, 2002, pp. 2-3)

One way of addressing the measurement and assessment of developing statistical literacy skills was suggested by Watson (1997). She suggested a three-tiered hierarchy, which includes the requirements (1) to understand basic statistical terminology, (2) to understand it in context, and (3) to question claims made without proper statistical justification. Although developed independently, the hierarchy displays similarities to the four roles of a critical reader, described for literacy more generally by Luke and Freebody (1997). Their roles include that of a code breaker to understand text, that of a text participant to make meaning, that of a text user to apply meaning in a social context, and that of a text analyst to think critically about claims made in text. Watson's hierarchy reflects whatever terminology is the focus of a particular task or investigation. Such terminology, for example "average," may include mathematical formulae or calculations. For the present study however non-numerical usage is the focus. Particular survey items are chosen to address the three tiers of the hierarchy. As an example Watson and Moritz (2000) considered questions asking for the definition of "sample" (Tier 1), asking for a choice of sample in the context of purchasing a car (Tier 2), and asking for judgments about claims made for samples in social contexts in the media (Tier 3). For each of the tiers they observed increasing levels of outcome reflecting the complexity and appropriateness of responses. The tiers themselves, however, were not strictly hierarchical and elsewhere it has been observed that the ability to produce sophisticated definitions is not likely to appear until students are also able to demonstrate the beginning of critical thinking with respect to tasks in applied contexts (Watson and Callingham, 2003). Watson and Kelly (2003) considered a larger data set that included the students in the present study and examined student understanding of the terms "sample," "random," and "variation." Improvements varied for different terms and different grades both after a unit of work and after a two-year period. Watson and Kelly (2005) also considered a set of questions related to sampling and sampling bias for the larger data set, again with varying results but with an indication of a dip in performance at the Grade 7 level compared to Grade 5 and Grade 9.

THE PRESENT STUDY

The present study is based on the performance of Grade 7 and 9 students on a set of items requiring statistical literacy skills that are not computationally based. Each question fits the description of one of Watson's (1997) hierarchical tiers and the following research questions are considered.

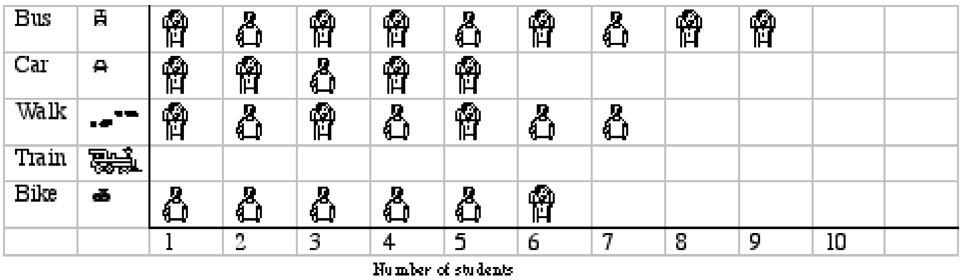
1. At the middle school level what is the base line performance of students in terms of (i) the statistical terminology, (ii) the terminology in context, and (iii) the critical questioning skills in context?
2. For these middle school students, how does their performance in these areas change six weeks after a unit of work on chance and data emphasizing variation?

METHOD

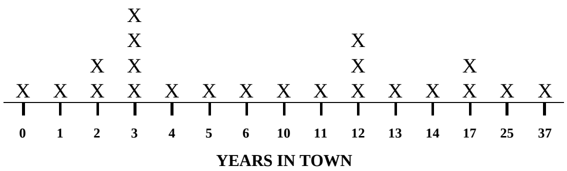
- *Sample*. Ninety students in each of Grade 7 (ages 12-13 years) and Grade 9 (ages 14-15 years) from two government schools in the Australian state of Tasmania completed pre and post tests surrounding a unit of work prescribed by their usual classroom teacher. Details of the procedures are found in Watson and Kelly (2002).
- *Items*. The items used in this analysis are shown in Figure 1, where they are labeled according to whether they are related to terminology (TERM), terminology in context (CONTEXT), or critical questioning in context (CRIT). Three items address terminology, asking for explanations of the meanings of the terms "sample," "random," and "variation." Eight items address terminology in context and relate to sampling (2) or to interpretation of graphical representations (6). Nine items require critical thinking in the context presented: sampling (7) and graphical interpretation (2).
- *Analysis*. All items were hierarchically coded using rubrics documented as part of a larger study of the initial data set by Watson, Kelly, Callingham, and Shaughnessy (2003). The ranges of code values for each item are given in Figure 1 in square brackets with the items. For the present study, totals for each of the three scales and the overall total were calculated, with means and standard deviations used for comparisons with paired *t*-tests. Both *p*-values and Cohen's (1969) effect sizes are reported in Table 1.

TERM1[0-3] a) What does "sample" mean? b) Give an example of a "sample."
TERM2[0-3] a) What does "random" mean? b) Give an example of something that happens in a "random" way.
TERM3[0-3] a) What does "variation" mean? b) Use the word "variation" in a sentence. c) Give an example of something that "varies."

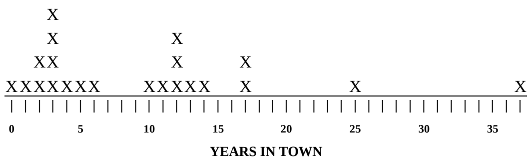
How children get to school one day



How many children walk to school? How many more children come by bus than by car?
CONTEXT1[0-1] Would the graph look the same everyday? Why or why not?
CONTEXT2[0-3] A new student came to school by car. Is the new student a boy or a girl? How do you know?
CONTEXT3[0-2] What does the row with the Train tell about how the children get to school?
CONTEXT4[0-5] Tom is not at school today. How do you think he will get to school tomorrow? Why?
A class of students recorded the number of years their families had lived in their town. Here are two graphs that students drew to tell the story.



CONTEXT5[0-3] What can you tell by looking at Graph 1?
CONTEXT6[0-3] What can you tell by looking at Graph 2?



CRIT1[0-3] Which of these graphs tells the story better? Why?
CONTEXT7[0-3] A class wanted to raise money for their school trip to Movieworld on the Gold Coast. They could raise money by selling raffle tickets for a Playstation 2. But before they decided to have a raffle they wanted to estimate how many students in their whole school would buy a ticket. So they decided to do a survey to find out first. The school has 600 students in grades 1-6 with 100 students in each grade. How many students would you survey and how would you choose them? Why?
CRIT2[0-3] *Shannon* got the names of all 600 children in the school and put them in a hat, and then pulled out 60 of them. What do you think of Shannon's survey? ☐ GOOD ☐ BAD ☐ NOT SURE Why?
CRIT3[0-3] *Jake* asked 10 children at an after-school meeting of the computer games club. What do you think of Jake's survey? ☐ GOOD ☐ BAD ☐ NOT SURE Why?
CRIT4[0-3] *Adam* asked all of the 100 children in Grade 1. What do you think of Adam's survey? ☐ GOOD ☐ BAD ☐ NOT SURE Why?
CRIT5[0-3] *Raffi* surveyed 60 of his friends. What do you think of Raffi's survey? ☐ GOOD ☐ BAD ☐ NOT SURE Why?
CRIT6[0-3] *Claire* set up a booth outside of the tuck shop. Anyone who wanted to stop and fill out a survey could. She stopped collecting surveys when she got 60 kids to complete them. What do you think of Claire's survey? ☐ GOOD ☐ BAD ☐ NOT SURE Why?
CRIT7[0-2] Who do you think has the best survey method? Why?

A primary school had a sports day where every child could chose a sport to play. Here is what they chose.

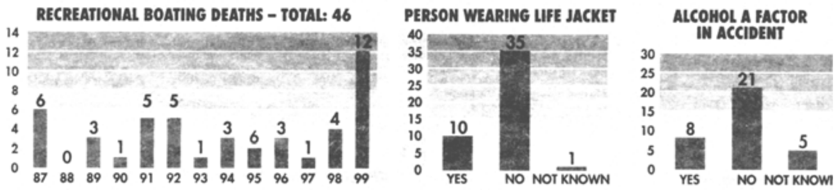
	Netball	*Soccer*	*Tennis*	*Swimming*	*Total*
BOYS	0	20	20	10	50
GIRLS	40	10	15	10	75

How many girls chose Tennis? What was the most popular sport for girls? What was the most popular sport for boys?

How many children were at the sports day?

CONTEXT8[0-4] The teacher wanted to choose four children to lead the closing parade. Suggest two fair ways she could have chosen them.

BOATIES' SAFETY FAILURE



These graphs were part of a newspaper story reporting on boating deaths in Tasmania.

CRIT8[0-2] Comment on any unusual features of the graphs.

Decriminalize drug use: poll

SOME 96 percent of callers to youth radio station Triple J have said marijuana use should be decriminalized in Australia. The phone-in listener poll, which closed yesterday, showed 9924 - out of the 10,000-plus callers - favoured decriminalisation, the station said. Only 389 believed possession of the drug should remain a criminal offence. Many callers stressed they did not smoke marijuana but still believed in decriminalizing its use, a Triple J statement said.

What was the sample size in this article?

CRIT9[0-2] Is the sample reported here a reliable way of finding out public support for the decriminalisation of marijuana? Why or why not?

Figure 1: Statistical literacy items

RESULTS

The statistics for both research questions are found in Table 1. The mean values in the second column are from the pretests taken by the students. For TERM and CONTEXT there is little difference between Grades 7 and 9, but for CRIT, there is an indication of better initial performance for the Grade 9 students. Overall the performances do not represent very well developed statistical literacy skills as the maximum possible score for TERM is 9, for CONTEXT is 24, and for CRIT is 24. The changes six weeks after completion of the units of work, represented by means in the third column, are all positive but although all are statistically significant, the effect sizes using Cohen's (1969) ranges are only small (0.15-0.44) or medium (0.45-0.74). Again this is considered a disappointing, although perhaps not unrealistic result.

Table 1: Comparison of pre and post means by grade

	Pre $\bar{X}$ (SD)	Post $\bar{X}$ (SD)	Difference	*t*	*p*	Effect
TERMINOLOGY						
Grade 7	2.47 (2.16)	3.34 (2.14)	0.88	3.83	<.0002	0.41
Grade 9	2.71 (2.26)	3.42 (2.51)	0.71	2.19	.016	0.30
CONTEXT						
Grade 7	10.56 (4.13)	12.41 (4.48)	1.85	4.39	<.0001	0.43
Grade 9	10.91 (4.20)	12.43 (4.63)	1.52	3.08	.0014	0.34
CRITICAL						
Grade 7	6.51 (5.39)	9.52 (6.30)	3.01	5.12	<.0001	0.51
Grade 9	8.89 (5.19)	10.19 (6.08)	1.30	1.99	.025	0.23
TOTAL						
Grade 7	19.53 (9.92)	25.28 (11.63)	5.75	5.68	<.0001	0.53
Grade 9	22.51 (9.82)	26.04 (11.42)	3.53	2.96	.002	0.33

Although overall the results are somewhat disappointing there are some examples of improved individual performance for particular items that illustrate the goals for statistical literacy as part of the wider chance and data curriculum. Several of the questions associated with the CONTEXT and CRIT variables are considered in this light.

In CONTEXT2 students were asked whether a new student who came to school by car was a boy or girl and how they decided. The increasing codes recognized greater involvement with the pictograph and at the highest code (3) making a prediction based on frequency, perhaps including uncertainty. A Grade 7 student moved from "Boy, because it goes in a pattern" at Code 1, observing the pattern in the graph, to "Girl, because there is more girls going in car than boys," at Code 3, observing the frequency in the graph. A Grade 9 student changed from "Boy, I'm guessing" at Code 0 to "Most likely a girl. According to the graph up there, girls like getting a car lift more than boys," at Code 3, also acknowledging the uncertainty involved.

For the CONTEXT4 question, where students were asked how Tom who was not at school today will come to school tomorrow, many students did not respond on the pretest, being allocated Code 0. One of these Grade 7 students, later on the post test used the graph in the response but not in a frequency sense, "Train. Because there is no one on the train row and if there wasn't no one the train wouldn't take the children to school and it wouldn't be on the chart" (Code 3). A Grade 9 student provided an imaginative response on the pretest at Code 1, "Car. He doesn't feel well," but improved to a frequency-based Code 4 response later, "Bus. More catch buses." A Grade 9 student was one of the few acknowledging uncertainty (Code 5) in both, e.g., predicting, "Probably by bus. Because one third of the children caught it today."

CONTEXT7 was about choosing a sample from a school of 600 students with 100 in each of the Grades 1 to 6. Sample size and method of selection were used as criteria for coding. Although the sample size was too large, a Grade 7 student moved from a Code 1 response of choosing specific people, "Around 300 - I would choose the people who don't have video games, and the computer nerds," to a Code 3 response where the method was appropriate, "300 people and I would choose them by random." Another Grade 7 moved from Code 1 of using the entire population, "I would survey all of them to find out exactly how many people would want them," to Code 3, "I would survey 10 from each grade; 5 boys and 5 girls. I would randomly pick them." This was one of the items where Grade 7 students showed over half of a code improvement in the mean score. CONTEXT 8 attracted similar responses but on this item it was the Grade 9 students whose mean increased by half a code level on the post test.

One of the CRIT questions, CRIT1, asked which of the two stacked dot plot better told the story of how long families had lived in a town. A Grade 7 student on the pretest chose Graph 1 as better after apparently misinterpreting the information in the graph, "Because they surveyed more people" (Code 0) but later chose Graph 2 because "It is more detailed" (Code 3). Another Grade 7 chose Graph 1, "Because more people lived in all the years" initially but later changed to Graph 2, "Cause they are spread out more" (Code 3). Some choices of Graph 1 were accompanied by observations of the layout of the graph, as for the Grade 9 student who wrote, "Because it's more clear and goes up in 1's not 5's" (Code 1). This student later chose Graph 2, "Because it's more spread out and it's more realistic to graph 1" (Code 3).

The item CRIT8 was assessed with respect to finding the errors in the graphs in the newspaper article: Code 0 for non-responses, incorrect or irrelevant observations, or reflections on the messages contained in the graphs; Code 1 for comments on the general shape of or numerical information in the graphs; and Code 2 for the actual misrepresentations or missing data in the graphs. The assessment was based on statistical appropriateness not strict correctness as many of the responses could not be judged correct or incorrect. A Grade 7 student who initially observed "The first one is set out right" (Code 0), later made two observations of the story in the graphs: "Not many people are wearing life jackets. There have been lots of deaths" (Code 1). A Grade 9 student who made no comment on the initial survey, in the second included both an implication and an error: "In 1998, 1990, 1993 and 1997, the government must have put their foot down on the rules. In 1995 the number says 6, but the line is on 3" (Code 2). A few students progressed from Code 1 to Code 2 after the instruction, as the following Grade 9: "Something big must've happened in '99 to get 12 deaths. Alcohol isn't a very big factor in boating deaths" (Code 1); "Most people died in 99. Most people didn't wear life jackets and most people weren't drinking. The total deaths in graph 1 actually added up to 48" (Code 2).

For the CRIT9 question about a phone-in survey on the use of marijuana, Code 1 responses generally said yes it was a good survey or no for an unintelligible reason. Some students who did not answer initially, provided a Code 2 reason based on sampling method in the second survey. From Grade 7, "No. Because they only interviewed JJJ callers" and from Grade 9, "No. If you thought it should stay illegal you wouldn't ring up" or "No. Because only young listeners would have phoned up." A few students changed from "yes" to "no" between the surveys, as did the following Grade 9, “Yes. A whole majority of people would have called in” and “No. Only the people listening will be able to vote.”

The three TERM items were discussed individually with examples of improved performance by Watson and Kelly (2003). Scores were not combined there but it was observed that only about 20% of variation was shared between them in pairs.

DISCUSSION

The changes in the present study, although encouraging, fall well short of educational expectations. A more concerted effort to get students to write words of explanation and think through the critical aspects of tasks is required. Given the attitude of many students and some teachers to the development of literacy skills in the mathematics classroom, it is difficult to see the situation changing soon. It may be that new curriculum initiatives that stress the integration of disciplines within broadly based educational goals, related to “providing a sustainable environment” or “living democratically,” may provide contexts for using statistical literacy skills across the curriculum. This will be useful especially in contexts where description and report writing are expectations of students and may enhance the statistical literacy skills themselves.

ACKNOWLEDGEMENT

The data collection and initial analysis for the present study were funded by australian research council grants nos. A00000716 and DP0208607.

REFERENCES

Cohen, J. (1969). *Statistical Power Analysis for the Behavioural Sciences*. New York: Academic Press.

Gal, I. (2002). Adults’ statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, 70, 1-51.

Luke, A. and Freebody, P. (1997). Shaping the social practices of reading. In S. Musprati, A. Luke, and P. Freebody (Eds.), *Constructing Critical Literacies: Teaching and Learning Textual Practice*, (pp. 185-225). St. Leonards, NSW: Allen and Unwin.

Ministry of Education. (1992). *Mathematics in the New Zealand curriculum*. Wellington: Author.

National Council of Teachers of Mathematics. (1989). *Curriculum and Evaluation Standards for School Mathematics*. Reston, VA: Author.

Wallman, K. K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association*, 88(421), 1-8.

Watson, J. M. (1997). Assessing statistical literacy using the media. In I. Gal and J. B. Garfield (Eds.), *The Assessment Challenge in Statistics Education*, (pp. 107-121). Amsterdam: IOS Press and The International Statistical Institute.

Watson, J. M. and Callingham, R. A. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal*, 2(2), 3-46.

Watson, J. M., and Kelly, B. A. (2002). Variation as part of chance and data in grades 7 and 9. In B. Barton, K.C. Irwin, M. Pfannkuch, and M.O.J. Thomas (Eds.), *Mathematics Education in the South Pacific: Proceedings of the 26th Annual Conference of the Mathematics Education Research Group of Australasia*, Vol. 2, (pp. 682-289). Sydney: MERGA.

Watson, J. M. and Kelly, B. A. (2003). The vocabulary of statistical literacy. In *Educational Research, Risks, and Dilemmas: Proceedings of the Joint Conferences of the New Zealand Association for Research in Education and the Australian Association for Research in Education* [CD-ROM]. Auckland.

Watson, J. M. and Kelly, B. A. (2005). Cognition and instruction: Reasoning about bias in sampling. *Mathematics Education Research Journal*, 17(1), 24-57.

Watson, J.M., Kelly, B.A., Callingham, R.A., and Shaughnessy, J.M. (2003). The measurement of school students’ understanding of statistical variation. *International Journal of Mathematical Education in Science and Technology*, 34, 1-29.

Watson, J. M. and Moritz, J. B. (2000). Development of understanding of sampling for statistical literacy. *Journal of Mathematical Behavior*, 19, 109-136.