ASSESSING STATISTICAL UNDERSTANDING IN GRADES 3, 6 AND 9 USING A SHORT-ANSWER QUESTIONNAIRE

Jane M. Watson, Kevin F. Collis and Jonathan B. Moritz
University of Tasmania

This paper presents some results of a pilot study which devised a 20-item paper-and-pencil, short-answer/multiple-choice questionnaire to assess students' understanding of statistics and probability in Grades 3, 6 and 9. The items are presented, with discussion of response differences over the three grades, and the level and type of cognitive functioning associated with responses.

The movement to include chance and data as part of the mathematics curriculum in Australian schools has been gaining pace since the publication of *A National Statement on Mathematics for Australian Schools* by the Australian Education Council (AEC) in 1991. The preparation of the *National Statement* and subsequent decisions made by state education authorities have taken place, however, without any associated research concerning students' abilities to cope with the topics at the levels suggested. The results reported here are from the pilot stage of a project designed to (i) follow the implementation of the chance and data curriculum in Tasmania, (ii) suggest appropriate concepts for students of various ages, and (iii) develop assessment procedures appropriate for the content in the curriculum. It is hoped that by the end of the decade a model of student understanding will emerge which will assist educators and curriculum designers with the revision of current statements, guidelines, and profiles.

Item development

The initial stages of the project focused on devising and trialling items to be used over the life of the project with students in Grades 3, 6 and 9. Four instruments were devised and trialled in three Tasmanian schools. A 20-item short-answer/multiple-choice paper-and-pencil questionnaire was based on items from other research studies overseas (sources are given in Table 1). A second survey of 11 items was devised based on extracts from newspapers which include reference to chance and data concepts. A third instrument included four interview items which used concrete materials, and the fourth instrument included six interview protocols modelled on the *Collis-Romberg Mathematical Problem Solving Profiles* (1992). This report will focus on the first of these instruments.

The theoretical model used as a framework for developing instruments is the SOLO model based on the taxonomy developed by Biggs and Collis (1982) and later extended by them (Biggs & Collis, 1991; Collis & Biggs, 1991). This neo-Piagetian model incorporates a multimodal functioning component, acknowledging the continued development of earlier modes of functioning in association with later modes. Of interest here are the ikonic and concrete symbolic modes, these being associated with intuitive functioning and the symbolic learning which takes place in school, often based on concrete materials. For other areas of mathematics it has been possible to identify multimodal functioning that is an interaction between the ikonic and concrete symbolic modes (e.g., Watson, Campbell & Collis, 1993), and two learning cycles operating in a hierarchy described as unistructural, multistructural and relational within the concrete symbolic mode (e.g., Campbell, Watson & Collis, 1992; Watson, Collis & Campbell, in press).

Pilot procedure

The trialling of the 20-item paper-and-pencil questionnaire took place in local schools of similar socio-economic status to the schools to be used in the main study. The sample sizes were n = 26 for Grade 3, n = 26 for Grade 6, and n = 32 for Grade 9. The items from the questionnaire are presented in Figures 1 and 2 in a much more compact format than that given to students. Grade 3 students were presented the first ten questions on five A-4 pages while Grade 6 and 9 students were presented the 20 items on seven A-4 pages. Students were allowed 40-50 minutes to complete the questionnaire. Table 1 presents references to specific objectives in the *National Statement* (AEC, 1991) associated with each item, as well as the source of each item. Of course it was not possible to cover all aspects of the curriculum in the 20 items, which is why three other instruments were developed.

Figure 1. First Ten Items - Grades 3, 6 and 9

1. Write two things which **might** happen today.
2. Write two things which **will** happen today.
3. Write two things which **won't** happen today.
4. Finish these sentences.
   (a) Tomorrow, it is **impossible** that________ (b) Tomorrow, it is **certain** that________________
   (c) Tomorrow, it is **possible** that__________ (d) Something that happens by **chance** is___________
5. Every morning, James gets out on the left side of the bed. He says that this increases his chance of getting good marks. What do you think?
6. One day, Claire won Tattslotto with the numbers 1; 7; 13; 21; 22; 36. So she said she would always play the same group of numbers, because they were lucky. What do you think about this?
7. Consider rolling one six-sided die. Is it easier to throw
   ❑ (1) a one or ❑ (6) a six or ❑ (=) are both a one and a six equally easy to throw?
8. A mathematics class has 13 boys and 16 girls in it. Each pupil's name is written on a piece of paper. All the names are put in a hat. The teacher picks out one name without looking. Is it more likely that
   ❑ (b) the name is a boy or ❑ (g) the name is a girl or ❑ (=) are both a girl and a boy equally likely?
   Please explain your answer.
9. Box A and Box B are filled with red and blue marbles as follows:

Box A	Box B
6 red 4 blue	60 red 40 blue

   Each box is shaken. You want to get a blue marble, but you are only allowed to pick out one marble without looking. Which one should you choose?
   ❑ (A) Box A (with 6 red and 4 blue). ❑ (B) Box B (with 60 red and 40 blue). ❑ (=) It doesn't matter.
   Please explain your answer.
10. A primary school had a sports day where every child could choose a sport to play. Here is what they chose:

	Netball	Soccer	Tennis	Swimming
Girls	30	5	15	10
Boys	0	20	18	20

   (a) How many girls chose tennis? __________
   (b) How many boys chose netball? __________
   (c) How many children chose swimming? _____
   (d) In which sport were boys and girls most evenly divided? _______________
   (e) Were there more girls or more boys at the sports day? ____ How do you know? ____

Figure 2. Last Ten Items - Grades 6 and 9

11. A bottle of medicine has printed on it:

> ***WARNING: For applications to skin areas there is a 15% chance of getting a rash. If you get a rash, consult your doctor.***

What does this mean?
- ❑ (a) Don't use the medicine on your skin - there's a good chance of getting a rash.
- ❑ (b) For application to the skin, apply only 15% of the recommended dose.
- ❑ (c) If you get a rash, it will probably involve only 15% of the skin.
- ❑ (d) About 15 out of every 100 people who use this medicine get a rash.
- ❑ (e) There is hardly any chance of getting a rash using this medicine.

12. To get the average number of children per family in a town, a teacher counted the total number of children in the town, then divided by 50, the total number of families. The average number of children per family was 2.2. Tick which of these is certain to be true.
- ❑ (a) Half of the families in the town have more than 2 children.
- ❑ (b) More families in the town have 3 children than have 2 children.
- ❑ (c) There are a total of 110 children in the town.
- ❑ (d) There are 2.2 children in the town for every adult.
- ❑ (e) The most common number of children in a family is 2.
- ❑ (f) None of the above.

13. A small object was weighed on the same scale separately by nine students in a science class. The weights (in grams) recorded by each student are shown below.

6.3 6.0 6.0 15.3 6.1 6.3 6.2 6.15 6.3

The median of this set of data is
- ❑ (a) the most common value. ❑ (b) the middle value.
- ❑ (c) the most accurate value. ❑ (d) the average value.

So, the median value is ________________ grams.

14. Please estimate: (a) Out of 100 men, how many are left-handed.
(b) Out of 100 left-handed adults, how many are men.

15. Please estimate: (a) The probability that you will miss a whole week of school this year.
(b) The probability that you will get a cold this year.
(c) The probability that you will get a cold causing you to miss a whole week of school this year.

16. Please estimate: (a) The probability that a woman is a school teacher.
(b) The probability that a school teacher is a woman.

17. A farmer wants to know how many fish there are in his dam. He took out 200 fish and tagged each of them, with a coloured sign. He put the tagged fish back in the dam and let them get mixed with the others. On the second day, he took out 250 fish in a random manner, and found that 25 of them were tagged. Estimate how many fish are in the dam.

18. A health survey was conducted in a sample of 100 men in Tasmania of all ages and occupations. Please estimate:
- (a) How many of the 100 men have had one or more heart attacks.
- (b) How many of the 100 men are over 55 years old.
- (c) How many of the 100 men both are over 55 years old and have had one or more heart attacks.

19. Mr. Jones wants to buy a new car, either a Honda or a Toyota. He wants whichever car will break down the least. First he read in <u>Consumer Reports</u> that for 400 cars of each type, the Toyota had more break-downs than the Honda. Then he talked to three friends. Two were Toyota owners, who had no major break-downs. The other friend used to own a Honda, but it had lots of break-downs, so he sold it. He said he'd never buy another Honda.
Which car should Mr. Jones buy?
- ❑ (T) Mr. Jones should buy the Toyota, because his friend had so much trouble with his Honda, while his other friends had no trouble with their Toyotas.
- ❑ (H) He should buy the Honda, because the information about break-downs in <u>Consumer Reports</u> is based on many cases, not just one or two cases.
- ❑ (=) It doesn't matter which car he buys. Whichever type he gets, he could still be unlucky get stuck with a particular car that would need a lot of repairs.

20. Every year, Susan selects about 5 young actors for the drama team who perform brilliantly at audition. Unfortunately, most of these kids turn out to be no better than the rest. Why do you suppose that Susan usually finds that they don't turn out to be as brilliant as she first thought?
- ❑ (a) In her eagerness to find new talent, Susan may exaggerate the brilliance of the performances she sees at the audition.
- ❑ (b) The actors probably just made some nice acts at the audition that were much better than usual for them.
- ❑ (c) The actors probably coast on their talent alone without putting in the effort for a consistently excellent performance.
- ❑ (d) The actors who did so well at the audition may find that the others are jealous, and so they slack off.
- ❑ (e) The actors who did so well are likely to be students with other interests, so they don't put all their energies into acting after the audition.

Table 1. Questionnaire Sources

Item	*National Statement* reference	Source
1-3	Chance language (A1)	Adapted from Fischbein & Gazit (1984), p.4, Q.A1-3.
4	Chance language (A1)	Original, modelled on AEC (1991), p.166.
5	Likelihood of everyday experiences (B1)	Adapted from Fischbein & Gazit (1984), p.4, Q.B2.
6	Order events (A3), interpret predictions (B3)	Adapted from Fischbein & Gazit (1984), p.4, Q.B3.
7	Order events (A3)	Original, modelled on Varga (1983), p.76; Green (1983).
8	Describe outcomes (A2, A3)	Adapted from Green (1982), in Borovcnik & Bentz (1991), p.81.
9	Describe, order outcomes (A3, B2)	Adapted from Konold & Garfield (1992), Q.7.
10	Interpret information (A5, B5)	Original.
11	Likelihood of everyday experiences (B1, C1)	Adapted from Konold & Garfield (1992), Q.2.
12	Measures of location (C7) [Average]	Adapted from Konold & Garfield (1992), Q.22.
13	Measures of location (C7) [Median]	Adapted from Konold & Garfield (1992), Q.1.
14	Describe, order everyday experiences (B1, B2)	Original, modelled on Pollatsek et al. (1987), p.260, Q.5.
15	Language/likelihood of everyday experiences (B1, C1)	Original, modelled on Tversky & Kahneman (1983), p.309.
16	Describe, order events (B1, B2, C2)	Pollatsek et al. (1987), p.260, Q.3
17	Analyse events, estimate populations (C2, D7)	Adapted from Fischbein & Gazit (1984), p.4, Q.A7.
18	Analyse everyday events (B1, B2)	Adapted from Tversky & Kahneman (1983), p.309.
19	Evaluate info. from daily life (C5, C9) [Sample size]	Adapted from Konold & Garfield (1992), Q.11.
20	Evaluate information (C5, C9) [Regression to mean]	Adapted from Nisbett, Krantz, Jepson & Kunda (1983), p.354.

Summary data

The first four questions, focusing on language and personal experience of chance events, often could not be classified as correct or incorrect. Of particular interest for the first three questions was the possibility of classifying responses into three categories: personal, school-based and wider-world. Examples are given in Table 2. This classification and how frequencies in categories change over the three age groups are discussed in detail by Watson, Collis and Moritz (1993). It is interesting to note that students are aware of the different contexts in which it is possible to classify the certainty of events happening. Of the 84 students, only four chose the same type of response for all questions. Question 4d elicited two types of response: (i) a definition, such as "possible", "unexpected", or "a coincidence", and (ii) an event, such as rain, or winning a game. Whereas only 8% of Grade 3 students responded with a definition, this increased to 35% at Grade 6 and 34% at Grade 9.

Table 2. Examples of Categories for Chance Events

Personal	School-based	World
I will eat my lunch	Silent reading	Rain
Jane will come by my house	We'll have Chinese	Daffy Duck in Grade 3
I will breathe	I will have Science	The earth will continue spinning

Table 3 presents the percentage correct for each of the parts of questions from Question 5 for Grades 3, 6 and 9. The 16 questions produced 24 responses which could be classified as correct. As can be seen in Figures 1 and 2, Questions 8 and 9 asked for explanations, which may not have been correct even if the answers were; Question 10 had five parts; Question 13 asked for a numerical answer as well as a multiple choice response; Question 14 was assessed as correct if part (b) were given a higher probability than part (a) and a further correct score was recorded if the value for (b) fell between .45 and .55.

Included in Table 3 are hypothesized minimum levels of response in the concrete symbolic mode required for acceptable answers to the items. Correct answers may be given at higher levels.

The notation refers to Unistructural (U_1 or U_2), Multistructural (M_1 or M_2), or Relational (R_1 or R_2) responses in the first or second U-M-R cycle of the concrete symbolic mode (see Campbell, *et al.* (1992) for a more detailed explanation). Further analysis to confirm these levels will take place as part of the main study.

Table 3. Summary Results by Question

ITEM	GRADE 3	GRADE 6	GRADE 9
Q5 (U_1)	85%	88%	100%
Q6 (M_1)	62%	58%	63%
Q7 (R_1)	50%	73%	97%
Q8 (R_1/U_2)	46%	88%	75%
Q8E (R_1/U_2)	35%	81%	69%
Q9 (M_2)	23%	65%	75%
Q9E (M_2)	4%	42%	50%
Q10a (M_1)	88%	100%	100%
Q10b (M_1)	88%	96%	100%
Q10c (M_1)	88%	100%	100%
Q10d (R_1/U_2)	73%	92%	100%
Q10e (R_1/U_2)	69%	92%	97%
Avg. Score	7.12	9.77	10.25

ITEM	GRADE 3	GRADE 6	GRADE 9
Q11 (U_2)	-	50%	75%
Q12 (U_2)	-	23%	25%
Q13 (U_2)	-	31%	81%
Q13no. (M_2)	-	8%	31%
Q14 (U_2)	-	69%	75%
Q14no. (M_2)	-	38%	59%
Q15 (R_2)	-	58%	53%
Q16 (M_2)	-	31%	34%
Q17 (R_2)	-	8%	16%
Q18 (R_2)	-	42%	56%
Q19 (R_2)	-	23%	38%
Q20 (R_2)	-	35%	16%
Avg. Score	-	13.92	15.84

For the first 12 items in Table 3, a significant difference of average scores between Grades was found ($F_{(2,81)} = 27.37$, $p<0.001$). A contrast revealed that the Grade 3 average was significantly lower than the other Grades ($t_{(81)} = 7.25$, $p<0.001$). This appears to reflect a ceiling effect on performance for many items from Grade 6. For the total of 24 items in Table 3, a t-test indicated Grade 9 students scored significantly higher than Grade 6 students ($t_{(56)} = 2.80$, $p<0.01$).

For 20 of the 24 parts of questions, the percentage correct increased from Grade 3 to Grade 6 to Grade 9 or from Grade 6 to Grade 9. It is of interest to consider those items where improvement did not occur with age. For Item 6, there was little change over the three groups, with many students saying just "no", with no explanation. Responses similar to "no, the luck has been used up, so different numbers should be chosen", made by two, five and five students in Grades 3, 6 and 9 respectively, were classified as incorrect. For Question 8 the Grade 6 students out-performed the other two groups in selecting a girl's name as more likely and correctly justifying the response. For Question 15 the difference was not great, and for Question 20 the percentage correct was quite low for both groups. In no case was the performance of Grade 6 significantly better than Grade 9.

Detailed look at selected items

This section must be read with constant reference to Figures 1 and 2.

Question 5 is interesting in that an increasing percentage of students (4%, 27% and 44%, respectively, for each grade) included a psychological flavour in their answers. These generally indicated that if James believed getting out of bed on the left side were lucky, then it might indeed help him. These answers were counted as correct because of the implication that the writer did not necessarily believe in the claim made.

Item 9 requires an understanding of proportion for its solution and some Grade 3 students may not have had the language necessary to express their understanding. It is interesting to note that many students could not explain an apparently correct answer. It may also be of some concern to teachers that 25% of Grade 9 students in an above average class gave incorrect answers. Item 10 on reading information from tables was included as a base line question and most Grade 3 students took considerable time but obtained correct answers.

Of the incorrect answers to Question 11, Grade 9 students were evenly split between options (a) and (e), while 31% of Grade 6 students chose (e) with the others choosing (a) or (b). No one chose (c). All options were selected in Question 12 about the average. The most popular response was (e) (54% and 50% of responses respectively for Grades 6 and 9).

Most Grade 6 students had not heard of median (Question 13) and only two could find it in the data set given. Of the 81% of Grade 9 students who selected (b), only 38% (31% overall) could correctly determine the value. This would surprise some Grade 9 teachers.

Questions 14 and 16 were chosen to test for an understanding of conditional statements and probabilities. In each case the response was counted as correct if the response to part (b) was less than the response to part (a). It was of interest to see (following Pollatsek, *et al.*, 1987) if Question 14, involving estimates of actual numbers of people, would be easier than Question 16, involving probabilities. This did turn out to be the case for both Grade 6 (69% to 31%) and Grade 9 (75% to 34%). The second response for Item 14 judged the reasonableness of the response to part (b) which should have been about 50 or ½. For Question 14, 38% of Grade 6 and 59% of Grade 9 students approximated these values. While the number of incorrect responses to Question 14 were roughly split between part (a) = part (b) and part (a) > part (b), for Question 16, 46% of Grade 6 and 53% of Grade 9 students said that the probabilities of the two events were the same. This question was also asked in some interviews and it was found that many students believed that the two questions were the same.

Questions 15 and 18 were devised to investigate the conjunction fallacy which is shown if the probability of part (c) is chosen to be greater than the minimum of the probabilities for parts (a) and (b). Answers were hence judged as correct if probability part (c) was less than or equal to the minimum in parts (a) and (b). Nineteen per cent of Grade 6 and 9% of Grade 9 students did not answer the question. Two students based responses on the number of weeks in the year (52) or the number of weeks remaining from the date of the survey (31).

The answers for Question 17 ranged from 175 to 5000 and there were 23 different answers given. Many students showed no working and generally the understanding of ratio was very poor at both grade levels.

The responses to Question 19 showed that both Grade 6 and 9 students were more likely to believe that it does not matter which car is purchased (46% and 53%, respectively) rather than accepting the statistical information provided. No Grade 9 but 23% of Grade 6 students believed that the evidence of friends should be the main deciding factor.

Although the correct response to Item 20 was the modal response for Grade 6, Grade 9 students preferred (a) (31%) or (e) (34%). It would appear that the idea of regression to the mean expressed in part (b) is not as intuitively acceptable to these students as other reasons offered.

Multimodal functioning

Although there was less scope for multimodal functioning in the 20-item questionnaire than in the other instruments developed as part of the larger project, responses to some items showed evidence of an interaction between ikonic and concrete symbolic functioning, and also of pure ikonic functioning in some explanations considered incorrect from a mathematical viewpoint. The following explanations are from Grade 3 students who would be expected to use the ikonic mode freely.

For getting out of bed on the left side in Item 5:

> "[Yes] Because the opiset [opposite] to left is right and if he gets out of his bed on the left side he gets his work right."

For using the same Tattslotto numbers in Item 6:

> "I think it's cheeting [cheating] because she's copping [copying] the ticket she won with."

For choosing (b) boy in Item 8:

> "a boy because we always have boys."

For choosing box B in Item 9:

> "because I like blue and red and marbles."

An interaction of ikonic and concrete symbolic functioning is shown in responses to Item 5 which included comments about James' beliefs in contrast to the respondents' understandings of chance. In the second half of the questionnaire, evidence of multimodal functioning was strongest in Items 19 and 20, which required rather sophisticated statistical understanding for the selection of correct responses. About fifty percent of students indicated that luck was most significant in buying a car in Item 19, and as noted earlier, students preferred responses to Item 20 more in line with their intuition of social situations.

Discussion

It may be surprising to some that there is so little improvement in performance from Grade 6 to Grade 9 in these data compared with what might be expected over three years. The fact that pilot work was based in two schools only may mean that curriculum bias could have caused this result. This will be an interesting point to follow-up in the main study. Because of the slow introduction of chance and data into the school curriculum, however, others may not be surprised at the seeming lack of progress over the three years. The Tasmanian Chance and Data Guidelines were not published at the time the pilot work was carried out, although the *National Statement* had been distributed to all schools and chance and data workshops were available to some teachers throughout the state. As the overall research project is funded for three years it will be of great interest to see if improvement occurs for groups over that period of time.

Acknowledgments

This research was funded by the Australian Research Council, Grant Number A79231392.

References

Australian Education Council. (1991). *A national statement on mathematics for Australian schools*. Carlton, Vic. : Author.

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO Taxonomy*. New York: Academic Press.

Biggs, J. B., & Collis, K. F. (1991). Multimodal learning and the quality of intelligent behaviour. In H. A. H. Rowe (Ed.), *Intelligence: Reconceptualisation and measurement* (pp. 57-76). Hillsdale, N. J.: Lawrence Erlbaum.

Borovcnik, M., & Bentz, H. J. (1991). Empirical research in understanding probability. In R. Kapadia & M. Borovcnik (Eds.), *Chance encounters: Probability in education*, (pp. 73-105), Dordrecht: Kluwer.

Campbell, K. J., Watson, J. M., & Collis, K. F. (1992). Volume measurement and intellectual development. *Journal of Structural Learning and Intelligent Systems, 11*, 279-298.

Collis, K. F., & Biggs, J. B. (1991). Developmental determinants of qualitative aspects of school learning. In G. Evans (Ed.) *Learning and teaching cognitive skills* (pp. 185-207). Melbourne: Australian Council for Educational Research.

Collis, K. F., & Romberg, T. A. (1992). *Collis-Romberg mathematical problem-solving profiles*. Hawthorn, Vic. : Australian Council for Educational Research.

Fischbein, E., & Gazit, A. (1984). Does the teaching of probability improve probabilistic intuitions? *Educational Studies in Mathematics, 15*, 1-24.

Green, D. (1983). Shaking a six. *Mathematics in Schools, 12*(5), 29-32.

Konold, C., & Garfield, J. (1992). *Statistical reasoning assessment. Part 1: Intuitive thinking*. (Draft document.) Scientific Reasoning Research Institute, University of Massachusetts, Amherst, MA.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review, 90*, 339-363.

Pollatsek, A., Well, A., Konold, D., Hardiman, P., & Cobb, G. (1987). Understanding conditional probabilities. *Organizational Behaviour and Human Decision Processes, 40*, 255-269.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgement. *Psychological Review, 90*(4), 293-315.

Varga, T. (1983). Statistics in the curriculum for everybody - how young children and how their teachers react. In D. R. Grey, P. Holmes, V. Barnett, & G. M. Constable (Eds.), *Proceedings of the First International Conference on the Teaching Statistics Vol II* (pp. 71-80). Sheffield, U. K. : Teaching Statistics Trust.

Watson, J. M., Campbell, K. J., & Collis, K. F. (1993). Multimodal functioning in understanding fractions. *Journal of Mathematical Behaviour, 12*, 45-62.

Watson, J. M., Collis, K. F., & Campbell, K. J. (in press). Developmental structure in the understanding of common and decimal fractions. *Focus on Learning Problems in Mathematics*.

Watson, J. M., Collis, K. F., & Moritz, J. B. (1993, September). *Assessment of statistical understanding in Australian schools*. Paper presented at the Statistics '93 conference, Wollongong, NSW.