Numeracy Benchmarks for Years 3 and 5: What about Chance and Data?

Jane M. Watson
University of Tasmania

Since January 1997 there has been much debate within Australia on the type of numeracy benchmarks which should apply to children in Years 3 and 5. As well as debate on the breadth and depth of understanding, there has been difficulty in some areas establishing what children actually do know and can do. Although the data included in this report were not collected to answer questions related to numeracy benchmarking, they may help inform the debate about what children in Years 3 and 5 know and can do in the area of chance and data.

Numeracy benchmarks for Year 3 and 5 children in Australia have been under consideration since January 1997 under the auspices of the Task Force on Literacy and Numeracy set up by the Australian Ministerial Council on Education, Employment, Training and Youth Affairs. With the advice of academics from around the country and work of people employed by the Curriculum Corporation, the following definition of Numeracy was the agreed foundation from which the benchmarks would be developed.

> Numeracy is the effective use of mathematics to meet the general demands of life at school and at home, in paid work, and for participation in community and civic life. For the purposes of this project, the numeracy benchmarks ... will incorporate the development of students' understanding and competence with number and quantity (ie, measurement), shape and location, and the handling and interpretation of quantitative data. (K. McLean, personal communication, March 3, 1997)

The debate that has surrounded the development of the benchmarks has variously related to the breadth of application involved in using numeracy skills, the depth of conceptual understanding, the need for agreement among all states on each detail included, and the lack of data available to back up claims of what Year 3 and 5 students understand and can do. This paper seeks to address only one issue. It will examine responses from Year 3 and 5 students to nine items in the area of chance and data (Moritz, Watson, & Pereira-Mendoza, 1996; Watson, Collis & Moritz, 1994, 1995, 1997). It is hoped that such information can provide a starting point, both for further debate about numeracy benchmarks and for following the progress of students as chance and data take a more prominent place in the Australian mathematics curricula.

Method

The data reported here were not collected to answer specific questions related to numeracy benchmarking in Australia. They were collected to document the cognitive development of students across the years of schooling, initially in 1993 from Years 3, 6 and 9, and to evaluate the implementation of the chance and data curriculum in Tasmania from that year. To that end further data were collected from the 1993 cohort in 1995 and 1997, with new students introduced to the study each year. The sample reported here is made up of children selected from eight state primary schools in the seven regions of the Tasmanian government school system. The schools represented rural, urban and semi-urban areas and were either primary or district high schools. Either all Year 3 (and later Year 5) students from the school completed the survey, or the classes that participated were chosen by the school rather than the researchers. For the purpose of this report, data from 864 Year 3 students in 1993, 1995 and 1997 have been combined, as have data from 703 Year 5 students in 1995 and 1997.

The nine items for which data are presented in this report (see Figure 1), were part of a 20-item chance and data survey (Watson, 1994) administered to students in class over 45 minutes. Year 3 students answered the first ten items and hence time was not a factor in the quality of responses. Where necessary students were read the items but no further help was provided. Item 1 of the survey, which asked for events that were certain, impossible or possible is not discussed here due to space restrictions (see however, Watson, Collis & Moritz, 1993; Moritz et al., 1996).

Q2. If someone said you were "average", what would it mean?

Q3. What things happen in a "random" way?

Q4. If you were given a "sample", what would you have?

Q5. 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?

Q6. 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?

Q7. 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?

Please explain your answer.

Q8. 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.

Q9. 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 box 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.

Q10. 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 1. Items from the chance and data survey.

The previous analysis of the 1993 data set of which some of these data are a subset was based on the Structure of Observed Learning Outcomes (SOLO) model (Biggs & Collis, 1982; Collis & Biggs, 1991). Detailed discussion of the model and its application to topics in chance and data is found in Watson et al. (1994, 1995, 1997) and Moritz et al. (1996). The requirements of a benchmarking exercise are consistent with the reporting of unistructural-multistructural-relational (U-M-R) levels of structural performance and hence the same model will be used here. The levels of response are the following.

IK: These responses are in the ikonic mode which is prestructural to the concrete symbolic mode, usually associated with imagination or tautology.

U_1 or U_2: Unistructural responses in the concrete symbolic mode employ a single element of the mode associated with the task set. U_1 responses are in the first cycle of the concrete symbolic mode, while U_2 responses are in the second and represent the consolidation of a first cycle concept which may have been structurally complex in its construction there but later is employed as a single element in more complex arguments.

M_1 or M_2: Multistructural responses in the concrete symbolic mode employ several elements of the mode associated with the task set, usually in a sequence. The distinction between M_1 and M_2 is again associated with the first or second cycle of processing within the mode and the nature of the elements used.

R_1 or R_2: Relational responses in the concrete symbolic mode relate together the elements of the mode in a fashion which is structurally consistent with the expectations of the first or second cycle and the nature of the task set. In the first cycle R_1 response are likely to be associated with the construction of a concept, while in the second cycle they are likely to be associated with the application of the concept in a more complex problem solving situation.

For purposes of statistical comparison, response levels were given integer values: 0 for IK, 1 for U_1, 2 for M_1, 3 for R_1, 4 for U_2, 5 for M_2, and 6 for R_2.

Of the Year 3 students, 560 were in the Year 5 group two years later and hence 243 of the Year 5 students had not been surveyed previously. These students were compared for each item with those who had been surveyed previously and it was found that for only one item (Q2) was there some evidence ($p = .04$) that students in Year 5 repeating the item did better than those who had not seen it before. Hence there was no overall effect which could be attributed to recall of the survey items. Cross-cohort analyses showed an improved performance for Year 3 students in 1997 on items Q2, Q3, Q4, and the last two parts of Q10, while Year 3 students in 1995 performed less well on Q8. The large sample sizes in the study meant that these differences amounted to a maximum difference of .2 of a SOLO level. Hence it was felt that the overall picture of performance of students would not be unduly biased by combining the data as done here. While the issue of this sample representing children in the rest of Australia is not simple, it should be noted that most other states and territories were implementing a similar curriculum from 1993-4 onwards.

Results

Language of Statistical Understanding (Items Q2, Q3 and Q4)

By middle primary school, students have often been exposed to chance and statistical terms which may provide an important foundation for later understandings. The following examples illustrate levels of student understanding for the terms "average", "random" and "sample" (see Figure 1, Q2, Q3 and Q4), according to the categorisation reported by Moritz et al. (1996). Some responses did not offer any appropriate meaning or example of the term, confusing it with other words.

P: [Random]: People kidnap people and demand some money. [Year 5]

Some responses indicated the student had heard the word in conversation, by offering examples of usage of the word, sometimes apparently without appreciating the meaning. Such responses were classified as ikonic (IK) for Q2 which asked for meaning, as an example is not in the required mode for the task set. For Q3 and Q4, the phrasing of the question meant that appropriate examples were classified as unistructural (U_1). Common examples for "average" included height and intelligence. Examples for "random" included natural unpredictable phenomena, a specific humanly constructed event or process, or games and competitions. For "sample", examples were foods and products, and natural samples in science or social science contexts.

IK [Average]: You are the average height of tallness. [Year 5]
U_1 [Random]: Wind. Rain. [Year 5]
U_1 [Random]: Like random breath test. [Year 3]
U_1 [Random]: Tattslotto balls. [Year 5]
U_1 [Sample]: Like a sample of urine. [Year 3]

Some responses used common, natural language to offer a single description related to the term used, and were also classified at the U_1 level.

U_1 [Average]: You are o.k. [Year 3]
U_1 [Average]: Normal. [Year 3]
U_1 [Average]: I am the same as everyone else. [Year 5]
U_1 [Random]: In no particular order. [Year 5]
U_1 [Random]: Not planned things. [Year 5]
U_1 [Sample]: A small something. [Year 5]
U_1 [Sample]: A demonstration. [Year 3]
U_1 [Sample]: Something the same as something else. [Year 5]

At the multistructural level (M_1), responses elaborated beyond the single general idea to describe multiple facets of the term. For "average", responses described aspects of how the average relates to a data set or to a method for obtaining the average from the data set. For "random", responses provided multiple examples from different contexts or an example with a simple defining characteristic. For "sample", responses added a complement to the U_1 concept, sometimes incorporating a concrete example.

M_1 [Average]: That I'm doing the same as most people in the class. [Year 5]
M_1 [Average]: That you weren't smart and you weren't dumb you're in between. [Year 5]
M_1 [Random]: Winning a lottery, eclipses. [Year 5]
M_1 [Random]: The weather is random, we cannot control it. [Year 5]
M_1 [Sample]: I'd have, if it was for clothes, a small piece of material. [Year 5]

For "average" and "sample" at the M_1 level the realisation of the part-whole nature of these concepts began to emerge. Compare for example, the third U_1 [Average] response with the first M_1 [Average] response.

Responses at the relational level (R_1) were rarely observed in responses by primary school students to these questions. This level of response demonstrates the acquired concrete symbolic concept of "average" as a representative measure of a data set, "sample" as a small part representing a whole, and "random" as unpredictable or uncontrolled with respect to natural events, while also possibly meaning unbiased or equally likely when concerning selection in games, competitions, or surveys.

R_1 [Sample]: A bit of something to show you what the whole thing is like. [Year 5]

The percentages of responses at each SOLO level are shown in Table 1, where it can be seen that by Year 5, 87% could at least give an example of an average and 76% had some appropriate idea about the word "sample", while only 35% were aware of the idea of "random". The corresponding percentages for Year 3 were 54%, 52%, and 9%.

Table 1
Percentage of Responses by Year to Items about Statistical Language

SOLO Level	Q2 "Average" Year 3	Q2 "Average" Year 5	Q3 "Random" Year 3	Q3 "Random" Year 5	Q4 "Sample" Year 3	Q4 "Sample" Year 5
NR/ P- Irrelevant	46	13	91	65	48	24
IK/U_1 - Example	17	10	5	16	17	15
U_1 - Concept	32	56	3	14	15	17
M_1	5	22	1	5	17	36
R_1	-	-	-	-	3	8
N	864	703	864	703	864	703

Beliefs about Luck (Items Q5 and Q6)

Students often hold beliefs about luck from an early age, and the difficulties of removing deeply held intuitions are commonly reported in the statistics education research literature. Levels of response to the items about luck (see Figure 1, Q5 and Q6), described by Watson et al. (1995), range from belief in luck, to uses of common phrases to express beliefs in negative recency of outcomes or equal likelihood. Some responses expressed agreement with belief about luck. These were classified as ikonic, i.e., prestructural to a concrete symbolic construction of situations involving chance.

IK [Bed]: It would help him get a good mark. [Year 3]

IK [Tattslotto]: They probably are lucky numbers. [Year 5]

At the U_1 level, responses offered disagreement to belief in luck.

U_1 [Bed]: He's just being superstitious. [Year 5]

U_1 [Tattslotto]: No I don't think they are lucky numbers. [Year 3]

U_1 [Tattslotto]: Wrong, lightning doesn't strike in the same place twice. [Year 3]

To the Tattslotto item, some responses rejected the idea of lucky numbers, and attempted to express the chance element involved, but did not clearly express equality of chance; these were classified as transitional U-M_1 responses.

U-M_1 [Tattslotto]: It was just lucky and they aren't lucky numbers. [Year 3]

U-M_1 [Tattslotto]: No it is the luck of the draw. [Year 5]

At the M_1 level, responses expressed reasons for James' success other than the side of the bed, or claimed other numbers had the same chance for Claire.

M_1 [Bed]: I think it doesn't matter, to get good marks you should study. [Year 5]

M_1 [Bed]: It doesn't matter what side of bed you get out of, except if your bed is up against a wall. [Year 5]

M_1 [Tattslotto]: Every number has an equal chance of coming up. [Year 5]

M_1 [Tattslotto]: She still has the same chance as everyone else. [Year 5]

Responses at the R_1 level rejected James' and Claire's belief in luck, saying that luck was not the mechanism, but using psychological reasons for justifying their action.

R_1 [Bed]: I think that if James thinks it increases his marks then he will do better work and get better marks, the bed doesn't have anything to do with it. [Grade 5]

R_1 [Tattslotto]: If she won with it one week and she didn't take them again and they came up she would be angry so I would take them again. [Year 5]

Table 2
Percentage of Responses by Year to Items about Luck

SOLO Level	Q5 Lucky side of the bed Year 3	Q5 Lucky side of the bed Year 5	Q6 Tattslotto numbers Year 3	Q6 Tattslotto numbers Year 5
NR/Irrelevant	10	4	10	5
IK - Yes lucky (or "Yes")	17	10	16	12
U_1 - Not lucky (or "No")	69	75	67	62
U-M_1 - Luck of draw	-	-	7	20
M_1 - Same chance	5	10	<1	2
R_1 - Psychology	1	2	-	<1
N	864	703	864	703

Table 2 contains the results for Years 3 and 5 for these questions, where it can be seen that while students in these years do not offer structurally complex responses, over 80% of Year 5 and over 70% of Year 3 students reject beliefs associated with luck.

Understanding of Chance Measurement (Items Q7, Q8 and Q9)

Responses to the items about chance measurement (see Figure 1, Q7, Q8 and Q9) were categorised as expressing basic uncertainty, qualifying chance, and quantifying chance, as described by Watson et al. (1997). At the ikonic level, personal experiences or intuitions were used to make decisions about events with uncertain outcomes, often with idiosyncratic beliefs.

IK [Die]: (1), One is an unlucky number for some people in games. [Year 3]

IK [Names]: (b), Because boys are better than girls. [Year 3]

IK [Marbles]: (=), Because I like blue better. [Year 3]

At the U_1 level there was recognition of basic uncertainty in the outcome, often expressed as "anything can happen."

U_1 [Die]: (=), You never know what number you will throw. [Year 5]

U1 [Names]: (=), You don't know if it will be a boy or a girl. [Year 3]

U_1 [Marbles]: (=), Because you aren't looking and it could be either. [Year 5]

At the M_1 level chance was qualified in some way, such as indicating which outcome is more likely, or if the chances are the same for two outcomes. Some responses expressed conflicting ideas in qualifying chance.

M_1 [Die]: (=), It is just as easy to throw either number. [Year 5]

M_1 [Names]: (g), Because there's more girls in the hat. [Year 3]

M_1 [Names]: (=), It is 50-50 because there is only a difference of 3. [Year 5]

M_1 [Marbles]: (B), Box B has more blue marbles. [Year 3]

M_1 [Marbles]: (B), Because there are more marbles. [Year 5]

At the R_1 level, chance was measured numerically for simple settings. This level did not clearly account for the more complex setting involving two boxes in Q9.

R_1 [Die]: (=), It is a one in six chance for both of them. [Year 5]

R_1 [Marbles]: (A), Because in box A there are only two more red than blue but in box B there are 20 more. [Year 5]

R_1 [Marbles]: (=), Because there is more red in each box. [Year 3]

Functioning in the second cycle of the concrete symbolic mode was required to justify the correct response of (=) for Q9 with appropriate mathematical reasoning (note the inappropriate reasoning in responses above which chose (=)). At the U_2 level responses displayed a single idea consistent with the use of ratio, but without further justification or explicit reference to measurement.

U_2 [Marbles]: (=), You've got the same chance of getting a blue or a red, so any. [Year 5]

At the M_2 level, responses often compared numbers across boxes in an attempt to describe the relationship between the boxes.

M_2 [Marbles]: (=), If you pick A you would probably get a red and it's the same with box B except there's 10 x more. [Year 5]

At the R_2 level, responses needed to present a correct mathematical quantitative comparison, using ratios and/or percentages to measure chance within each box.

R_2 [Marbles]: (=), Because you have a 40% chance of getting a blue marble in each. [Grade 5]

Results for items Q7, Q8 and Q9 are shown in Table 3. Between 18% and 20% of Year 3 students and between 10% and 14% of Year 5 students gave responses to the multiple choice items but did not justify their responses with reasoning. These responses could not be allocated to SOLO levels. For Q7 the modal response for each year was associated with ikonic beliefs about dice. For Q8 however, in both groups over half of the students made sensible qualitative statements about the chances involved. For the more difficult Q9, only 11% of Year 3 and 21% of Year 5 students gave responses justifying the correct multiple choice selection which were in the second cycle of the concrete symbolic mode.

Table 3
Percentage of Responses by Year to Items about Chance Measurement

SOLO Level	Q7 Six-sided die Year 3	Q7 Six-sided die Year 5	Q8 Names in hat Year 3	Q8 Names in hat Year 5	Q9 Marbles in boxes Year 3	Q9 Marbles in boxes Year 5
NR	3	4	4	2	4	3
Incorrect choice	8	6	11	6	10	6
Correct choice only	10	8	7	4	10	7
IK	57	33	15	10	4	2
U_1	14	24	13	16	10	10
M_1	7	22	51	62	52	43
R_1	<1	4	-	-	7	6
U_2					3	12
M_2					1	2
R_2					-	<1
N	864	703	864	703	864	703

Understanding Data in a Table (Item Q10)

Q10 (see Figure 1) was intended to provide base-line information on data interpretation skills for the youngest students. Parts (a) and (b) were U_1 level questions for picking individual data values from a table. Part (c) was a straightforward M_1 question requiring the sum of numbers in two cells. Although an explanation was not sought for Part (d) and hence guessing was possible, it is likely that R_1 reasoning is required to compare girls and boys across each category. Part (e) was a more complex M_1 question which also asked for an explanation. The results for each part of the question are given in Table 4, with explanations for Part (e) in Table 5. Virtually all students performed at the M_1 level or higher on Q10, but considerably fewer students demonstrated correct calculations when totalling in Part (e). Many students gave a correct answer of "girls" by adding or counting but did not produce their totals. A few students in the "Add/count group" used a calculator. "Other reasons" included those choosing girls because they participated in every sport while boys did not.

Table 4
Percentage of Responses Correct to Item Q10 by Year

Correct	Year 3	Year 5
(a) 15	92	94
(b) 0	95	99
(c) 30	89	96
(d) Tennis	55	80
(e) Girls	71	81
N	864	703

Table 5
Percentage of Responses by Reason for Item Q10e by Year

Reason	Year 3	Year 5
60 and 58	15	10
60 or 58	9	3
Girls, Add/Count	35	54
Other, Add/Count	9	8
Other reasons	31	26
N	864	703

Discussion

The results of these analyses point to several useful findings for those who would describe the capabilities of Year 3 and 5 students in relation to chance and data. It is evident that nearly all students by Year 3 can successfully read straightforward information from tables. Most have a basic foundation for building a more sophisticated understanding of average and sample, and have understood that belief in luck is contrary to the operation of chance events. It also appears that chance situations involving dice present more difficulty for interpretation than chance situations involving simple sampling, such as from a hat. This may be related to children's experiences playing games with dice, for example when they need a certain number to start. It was also found that performance improved between Years 3 and 5 for each of the items used (noting the ceiling effect for Q10). It would appear that students in Years 3 and 5 are building the foundations of understanding to handle the learning experiences associated with the objective expressed in *A National Statement on Mathematics for Australian Schools* (Australian Education Council, 1991).

> A sound grasp of concepts in the areas of chance, data handling and statistical inference is critical for the levels of numeracy appropriate for informed participation in society today. (p. 163)

The questions in this study were designed to be relevant to older students as well as Years 3 and 5, in order to trace the development across all of the years of schooling. Hence many trivial questions were not included. Asking other questions of a more basic nature, relating to simple outcomes from chance events, sampling and interpretations of graphs, would have provided different information which might be used for benchmarking. As part of the larger study which collected these data, interviews were conducted with over 100 students using such questions, and this information will further inform the understanding of children's development in Years 3 and 5. As well it will be possible to describe longitudinal development over time (e.g., Watson & Moritz, 1998). As the curriculum becomes more firmly established, future research involving teaching experiments will reveal how much more students can achieve.

Acknowledgment

This research was funded by the Australian Research Council, Grants No. A79231392 and No. A79800950. The author thanks Jonathan B. Moritz for assistance in writing this paper.

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.

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). Australian Council for Educational Research, Melbourne.

Department of Education and the Arts, Tasmania. (1993). *Mathematics Guidelines K-8*. Hobart: Curriculum Services.

Moritz, J. B., Watson, J. M., & Pereira-Mendoza, L. (1996, November). *The language of statistical understanding: An investigation in two countries*. Paper presented at the Joint Conference of the Educational Research Association and the Australian Association for Research in Education, Singapore.

Watson, J. M. (1994). Instruments to assess statistical concepts in the school curriculum. In National Organizing Committee (Ed.), *Proceedings of the Fourth International Conference on Teaching Statistics. Volume 1* (pp. 73-80). Rabat, Morocco: National Institute of Statistics and Applied Economics.

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.

Watson, J. M., Collis, K. F., & Moritz, J. B. (1994). Assessing statistical understanding in Grades 3, 6 and 9 using a short answer questionnaire. In G. Bell, B. Wright, N. Leeson, & G. Geake (Eds.), *Challenges in mathematics education: Constraints on construction* (pp. 675-682). Lismore, NSW: MERGA.

Watson, J. M., Collis, K. F., & Moritz, J. B. (1995). Children's understanding of luck. In B. Atweh & S. Flavel (Eds.), *Proceedings of the Eighteenth Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 550-556). Darwin, NT: MERGA.

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. (1998). *The longitudinal development of chance measurement.* Manuscript submitted for review.