NATIONAL TESTING OF PROBABILITY IN YEARS 3, 5, 7, & 9 IN AUSTRALIA: A CRITICAL ANALYSIS

STEVEN NISBET
Faculty of Education, Griffith University
s.nisbet@griffith.edu.au

This paper provides a critical analysis of the probability questions in the 2009 & 2010 NAPLAN numeracy tests for Years 3, 5, 7, and 9 from a number of perspectives. The analysis revealed that probability is under-represented in recent NAPLAN tests, with only one probability item included in each year level in the 2010 test. The items reveal a limited focus regarding type of task and response method. There is poor coverage of the National Statements on Learning in Mathematics, and weak alignment with the constructs in the *Probabilistic Reasoning Framework* (Jones et al, 1997).

Introduction

In Australia, schools now have to provide evidence to the community that their students perform to 'acceptable standards'. In particular, in 1997 the Australian Federal Government adopted compulsory national testing of literacy and numeracy to (a) identify students at risk, (b) conduct intervention programs, (c) assess all students against national benchmarks, and (d) introduce a national numeracy reporting system (Department of Education, Training, & Youth Affairs, 2000). Since 2008, the writing of the numeracy tests has been done nationally, and the test items have been based on the *National Statements of Learning for Mathematics* (Curriculum Corporation, 2006). The tests are currently conducted in all schools in May, yearly.

The introduction of state and national testing has not been supported widely in the educational community in Australia. Nisbet & Grimbeek (2004) found that many Queensland teachers had very negative attitudes to state tests. They did not believe the tests were valid, and hence the results of the tests did not influence their teaching practices. Nor did the teachers use the test results to any great extent to inform their planning, apart from some identifying gaps in their schools' programs. National testing has limited validity considering (i) the accepted definitions and purposes of assessment and (ii) the nature of large-scale tests. First, assessment can be defined as the comprehensive accounting of a student's knowledge and a means to achieve educational goals, and not the end of an educational experience (Webb, 1993). The purposes of assessment according to Clarke, Clarke, and Lovitt (1990) are to: (i) improve instruction by identifying sources of error; (ii) improve instruction by identifying instructional strategies that are successful; (iii) inform the learner of strengths and weaknesses;

(iv) inform subsequent teachers of students' abilities; and (v) inform parents of their child's progress. The National Assessment Program—Literacy and Numeracy [NAPLAN] can only provide a 'snapshot' of students' achievements and satisfy these purposes to a limited extent.

Second, large-scale numeracy testing brings with it a number of logistical limitations. For instance, the methods of students' responses adopted in the NAPLAN tests are limited to just two types: (i) multiple choice ("Colour in the bubble"); and (ii) single numerical answers ("Write the number in the box"); that can be marked by computer scanning. These limit the questions and tasks that can be included in the tests. It is commonly accepted that through assessment, teachers communicate to students which activities and learning outcomes are valued, so that assessment should be comprehensive and give recognition to all valued learning experiences (Clarke, Clarke, & Lovitt, 1990). Thus national testing brings limitations including a bias towards mechanical processes, and away from problem solving and creativity.

It is known that items in a related area, statistics, in traditional pencil-and-paper mathematics tests have shortcomings; in that they test skills in isolation from the problem context, they do not test whether or not students understand how statistical measures are interpreted, and they fail to assess students' ability to communicate using statistical language (Garfield, 1993). More recent research by Nisbet revealed that the statistics items in NAPLAN were limited in type of tasks, and related to just one aspect of the requirements stated in the *National Statement of Learning: Mathematics* (Curriculum Corporation, 2006), namely, analysing data, and ignored data collection, representation and interpretation. Similarly, most of the statistics items (94%) aligned with just one construct of the *Statistical Thinking Framework* (Jones et al, 2001), and disregarded the other three constructs.

National testing programs have tended to result in classroom assessment moving away from authentic formative practices and towards techniques closely aligned to the national test format (Stiggins, 1999). Teachers feel compelled to spend time preparing their students to master the skills included on the tests. It is against this background that this analysis has been conducted.

NAPLAN numeracy tests

The Australian national numeracy tests are based on a broad definition of numeracy, viz. "Numeracy is the effective use of mathematics to meet the general demands of life at school, at home, in paid work, and for participation in community and civic life" (MCEETYA Benchmarking Task Force, 1997, p. 4). Hence the tests cover the strands of mathematics which most people would meet in daily life; i.e., number, measurement, geometry, chance, and data. The Year 3 test usually contains 35 items, and the Year 5 test 40 items. There are two tests at the Year 7 level, one done without a calculator (32 items), and another where a calculator is allowed (also 32 items). Similarly, there are two tests at the Year 9 level, one with and one without calculators (each 31 items in 2009, and 32 items in 2010). The methods of response to the test items vary across items; some items are answered by multiple choice, in which students have to colour in with pencil one of the small 'bubbles' placed under four alternatives, and some items are open response, in which students write their numerical answers in a small box

(30mm x 12mm) provided on the page. Each paper has a set of three practice items to assist students become familiar with how to answer the questions.

Another feature of the tests is the use of 'link items', i.e., items that are common to two year levels. These are inserted to compare performance across grades levels. Seven of the 32 items in the Year 7 non-calculator test are linked to the Year 5 test, and another seven items in the Year 7 non-calculator test are linked to the Year 9 calculator test. Typically, the percentage of students who answer a link item correctly in Year 7 is greater than the percentage of students who answer it correctly in Year 5.

Analysis of probability items in 2009 and 2010 tests

This paper provides an analysis of probability items with respect to the following issues: number of items included, definition of numeracy and use of real-world contexts; type of stimulus material; type of task; alignment with the *National Statements of Learning* [SOLs] for Mathematics for Years 3, 5, 7, and 9 (Curriculum Corporation, 2006); and alignment with the *Probabilistic Reasoning Framework* (Jones et al, 1997). Joliffe (2005) argues that it is useful to compare assessment tasks with curriculum goals and teaching and learning frameworks "to check what dimensions are being assessed" (p. 328).

In comparison with other topics in the mathematics syllabus, probability is under-represented in the NAPLAN tests, with respect to number of items included. In 2010, there was only one Chance item in each year-level test (approximately 3% of the items). In 2009, there was only one Chance item in each of the Year 3 and Year 5 tests (also 3%), two items in the Year 7 test (6%), and four items in Year 9 (6.5%).

The definition of numeracy underlying the Year 3, 5, 7, & 9 numeracy tests was adopted by the Australian Ministerial Council on Education when national benchmarks were set in 1997, and reads as follows: 'Numeracy is the effective use of mathematics to meet the general demands of life at school, at home, in paid work, and for participation in community and civic life' (MCEETYA Benchmarking Task Force, 1997, p. 4). All chance questions in the 2009 and 2010 tests referred to various types of hands-on materials, namely, spinners, dice, coins, marbles, jelly beans, and polyhedral blocks. However, such contexts may not be authentic for some students, as the authenticity of the contexts depends on the students' varied personal backgrounds and experiences, and even the extent to which their teachers had included activities utilising such materials in their classroom mathematics programs. The stimulus material for each item was a statement about particular concrete materials, mainly spinners and jelly beans. Other materials were dice, coins, marbles and blocks. Six out of 12 of the items (50%) included a drawing of the materials (e.g., a spinner), and three out of the 12 items (25%) included a table of results from a Chance experiment.

The items required the students to do one of the following tasks: (i) identify an outcome which is impossible to occur; (ii) identify an outcome which is most likely to occur; (iii) identify the most likely set of results; (iv) identify a situation or concrete material that optimises a particular outcome; (v) calculate the probability of an event occurring based on a diagram of particular concrete materials (e.g. spinner); and (vi) calculate the probability of an event occurring based on a given table of results. Only one item (out of the 12 items in total) required the students to write an answer after performing a calculation. The other 11 questions were multiple-choice items, which allowed students to guess rather than think about the item. Research has shown that it is possible to get multiple-choice items correct without knowing much or doing any real thinking (Northwest Regional Educational Laboratory, 1988). A student may pick the correct response by either knowing or calculating the correct answer, making an informed guess, or just guessing wildly (Parkes, 2010).

The Year 3 Statement of Learning relating to probability (Curriculum Corporation, 2006) is as follows: '[Students] make simple statements, including predictions about likelihood, what is possible and what is not.' (p. 6). The test item given in 2009 asked the students to identify which colour marble was impossible to pick from a box of coloured marbles (with information given about numbers of each colour), so it aligns with the SOL. The item given in 2010 asked the students to identify which spinner from a group of four spinners (illustrated) was most likely to stop on 'white', so it aligns with the SOL also. Hence it is concluded that the Year 3 test items addressed the Year 3 SOL, albeit it in a limited way, with only one item included each year.

The Year 5 Statement of Learning (Curriculum Corporation, 2006, p. 10) refers to students:

- identifying and describing all possible outcomes for familiar chance events;
- making judgments about their likelihood;
- predicting whether some outcomes are more likely than others;
- using suitable language including most unlikely, never, probably;
- collecting data from experiments to justify or adjust these predictions; and
- distinguishing situations that involve equally-likely events from those that do not.

The probability item included in the Year 5 test in 2009 asked the students to identify "Which jar gives Annie the best chance of taking a black jelly bean?" so it aligns with the third dot point. The item in 2010 asked "On which number on a spinner was the arrow most likely to stop?", so it aligns with the third dot point also. It is concluded that the Year 5 test items assessed just one out of the six aspects of chance referred to in the Year 5 SOL. This seems unsatisfactory, but is not surprising considering that the tests contained only one item on probability each year.

The Year 7 Statement of Learning refers to students doing the following:

- comprehending that many events have different likelihoods of occurrence;
- making and interpreting empirical estimates of probabilities;
- comparing experimental data for simple chance events with theoretical probability obtained from proportions expressed as percentages, fractions or decimals; and
- distinguishing events that are equally likely from those that are not. (p. 7)

There were two probability items in the Year 7 test in 2009. The first asked the students "What is the chance that a jelly bean (drawn from a tin) is red?" so it aligns with the second dot point. The second item showed a spinner with unequal sectors and asked "Which table is most likely to show the results of spinning it 100 times?" It aligns with the first and fourth dot points. The only chance item in the 2010 test asked the students to identify which "On which number on the spinner is the arrow most likely to stop?" so it also aligns with the first and fourth dot point of the Year 7 SOL. Hence it is concluded that the Year 7 test items assessed only two out of four aspects (dot points) of the Year 7 SOL.

The Year 9 Statement of Learning refers to students doing the following:

- using a variety of sources, including surveys, data-bases, experiments and simulations to estimate probabilities associated with events;
- assigning or making estimates of probabilities based on personal experiences;
- specifying sample (event) spaces for single and straightforward compound events using a variety of suitable representations;
- determining corresponding probabilities by counting, measure and symmetry; and
- being familiar with the notion of equally likely events; and
- being familiar with the use of random event generators, including technology. (p. 11)

There were four probability items included in the Year 9 test in 2009. The first item asked the students "What the probability is of rolling a 3 on a standard six-sided die?" This aligns with the second dot point. The second item asked "What is the chance that a jelly bean (drawn from a tin) is red?", so it aligns with the fourth dot point. The third item showed a spinner with unequal sectors and asked students to indentify "Which table is most likely to show the results of spinning it 100 times?" It aligns with the fourth dot point of the Year 9 SOL. Item 4 said "Calculate the probability of getting 2 tails and 1 head in any order when a coin is tossed 3 times." It aligns with the third and fourth dot points. It is concluded that the Year 9 test items in 2009 assessed four out of six aspects of the Year 9 SOL. The only Chance item in the 2010 test required the students to calculate the probability of rolling a 2 on a non-regular hexahedron block, given the frequency results for it being thrown 1000 times, so it also aligns with the first and fifth dot point of the Year 9 SOL. Hence it is concluded that the Year 9 test item in 2010 assessed two out of six aspects of the Year 9 SOL. A summary of the aspects of the Year 3, 5, 7, & 9 SOLs covered by NAPLAN is shown in Table 1.

Table 1: Aspects of National Statements of Learning covered in NAPLAN tests by Year level.

Year of test	Year level			
	Year 3	Year 5	Year 7	Year 9
2009	1 out of 1 (100%)	1 out of 6 (16.7%)	2 out of 4 (50%)	4 out of 6 (66.7%)
2010	1 out of 1 (100%)	1 out of 6 (16.7%)	2 out of 4 (50%)	2 out of 6 (33.3%)

Analysis in relation to the *Probabilistic Reasoning Framework*

The *Probabilistic Reasoning Framework* (Jones, Langrall, Thornton, & Mogill, 1997) was developed after research and validation with elementary school pupils. Its purpose is to describe students' probabilistic reasoning skills. It consists of six constructs that are described at four levels (Level 1, Subjective; L2, Transitional' L3, Informal quantitative; & L4, Numerical). The construct *Sample Space* refers to students being able to list the outcomes of one- and two-stage experiments. *Experimental probability* involves recognising the role of the number of trials, and calculating probability values from data. *Theoretical probability* involves being able to predict least likely and most likely events in one- and two-stage experiments, and assign numerical values to probability situations. *Probability comparisons* involve reasoning to distinguish 'fair' and 'unfair' probability situations. *Conditional probability* relates to how probability values differ in replacement and non-replacement situations. Finally, the construct *Independence* refers to students being able to distinguish between independent and non-independent events, and understanding the independence of consecutive trials.

In the Year 3 tests, only two out of the six constructs were assessed (Sample Space in 2009 and Probability comparisons in 2010). In the Year 5 tests, only one construct was assessed (Probability comparisons in both years). In Year 7, two constructs were assessed in 2009 (Experimental & Theoretical probability), and only one construct (Probability comparisons) was assessed in 2010. See Table 2.

Table 2: Probability constructs tested and levels of thinking required according to year level.

	Test items and levels of thinking required				
Construct	Year 3	Year 5	Year 7	Year 9	Totals
Sample Space	2009, Q15: L2				1
Experimental Probability			2009, Q17: L4	2010, Q16: L4	2
Theoretical Probability			2009, Q12: L4	2009, Q2: L4 2009, Q 10: L4 2009, Q28: L4	4
Probability Comparisons	2010, Q7: L2	2009, Q29: L2 2010, Q 7: L2	2010, Q 3: L2	2009, Q 15: L4	5
Conditional Probability					0
Independence					0
Total items	2	2	3	5	12

Overall the majority of the test items (9 out of 12 or 75%) were limited to two of the six constructs (Theoretical probability and Probability comparisons). Two of the 12 items related to Experimental Probability, and one item related to Sample Space. Two constructs (Conditional Probability & Independence) were not tested at all over the two years.

Despite the inadequate coverage, the levels of thinking expected of students followed a reasonable progression from Year 3 to Year 9.

Conclusion

The probability questions in the 2009 and 2010 tests relate to the MCEETYA (1997) definition of numeracy to a limited extent only, because of the variable levels of authenticity of the item contexts. There is also the potential for language or cultural bias, as has been noted in research concerning multiple-choice questions (Parkes, 2010). Further, the probability items are limited in the type of task required, with most being multiple-choice items in which students select a spinner, choose an outcome, or pick a probability value. The multiple-choice items are fundamentally recognition tasks, where students identify the correct response (Parkes, 2010). The test results do not indicate whether the students knew the answer, or made an informed or blind guess (Gronlund & Linn, 1990).

This analysis reveals that the *National Statements of Learning* (Curriculum Corporation, 2006) relating to probability are poorly served by recent NAPLAN tests.

Probability deserves to have more than one item included in the tests each year, and the items should cover more aspects of probability in the statements than those covered in the 2009 and 2010 tests. The current situation promotes a limited view of probability that may be noted by teachers and students and lead to a down grading of the topic in class.

The probability items in the 2009 and 2010 tests relate to a minority of the constructs defined by the *Probabilistic Reasoning Framework* (Jones et al, 1997). With no items on sample space, independence, and conditional probability the tests imply that these constructs are not important, and teachers may omit reference to them in their teaching. As noted by Sadler (1998), when narrow tests define learning, instruction often gets reduced to "drill and kill"— practice on questions that look like the test. If schools place emphasis on the content and style of items contained in NAPLAN tests, and teachers teach to the test (Stiggins, 1999), there will be an unfortunate narrowing of the enacted curriculum and students' knowledge and understanding of probability will suffer.

References

Clarke, D., Clarke. D., & Lovitt, C. (1990). Changes in mathematics teaching call for changes in assessment alternatives. In T. Cooney & C. Hirsch (Eds.), *Teaching and learning mathematics in the 1990s, NCTM 1990 Yearbook* (pp. 118–129). Reston, VA: NCTM.

Curriculum Corporation (2006). *National statements of learning for mathematics.* Carlton, Vic: Author.

Department of Education, Training & Youth Affairs, (2000). *National literacy and numeracy plan.* Canberra, ACT: Author.

Garfield, J. (1993). An authentic assessment of students' statistical knowledge. In N. Webb & A. Coxford (Eds.), *Assessment in the Mathematics Classroom. NCTM 1993 Yearbook* (pp. 187–196). Reston, VA: National Council of Teachers of Mathematics.

Gronlund, N. E.,& Linn, R. *Measurement and Evaluation in Teaching.* (6th ed.) New York: Macmillan, 1990.

Jones, G., Langrall, C., Thornton, C. & Mogill, A. (1997). A framework for assessing and nurturing young children's thinking in probability. *Educational Studies in Mathematics, 32*, 101–125.

Jones, G., Langrall, C., Thornton, C., Mooney, E., Perry, B., Putt, I., & Nisbet, S. (2001). Using students' statistical thinking to inform instruction. *Journal of Mathematical Behavior, 20*, 109–144.

MCEETYA Benchmarking Task Force (1997). *National report on schooling in Australia.* Canberra, ACT: MCEETYA.

Northwest Regional Educational Laboratory (1998) *Measuring thinking in the classroom*. Oak Park, IL: Author.

Nisbet, S., & Grimbeek, P. (2004). Primary teachers' beliefs and practices with respect to compulsory numeracy testing. In I. Putt, R. Faragher & M. McLean (Eds.) *Mathematics Education for the Third Millennium.* Proceedings of the 27th Conference of the Mathematics Education Research Group of Australasia (pp. 1–12). Townsville, Qld: MERGA.

Parkes, J. *Field-tested learning assessment guide: Multiple-choice test.* Educational Psychology Program, University of New Mexico. Accessed 23rd December, 2010, from http://www.flaguide.org/cat/mutiplechoicetest/multiple_choice_test7.php

Sadler, P. (1998). Psychometric models of student conceptions in science. *Journal of Research in Science Teaching, 35*(3), 265–296.

Stiggins, RJ. (1999). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practices, 18*(1), 23–27.

Webb, N. (1993). Assessment for the mathematics classroom. In N. Webb & A. Coxford (Eds.) *Assessment in the Mathematics Classroom. NCTM 1993 Yearbook* (pp. 1–6). Reston, VA: National Council of Teachers of Mathematics.