

THE DEVELOPMENT OF STATISTICAL LITERACY AT SCHOOL

ROSEMARY CALLINGHAM
University of Tasmania
Rosemary.Callingham@utas.edu.au

JANE M. WATSON
University of Tasmania
Jane.Watson@utas.edu.au

ABSTRACT

Statistical literacy increasingly is considered an important outcome of schooling. There is little information, however, about appropriate expectations of students at different stages of schooling. Some progress towards this goal was made by Watson and Callingham (2005), who identified an empirical 6-level hierarchy of statistical literacy and the distribution of middle school students across the levels, using archived data from 1993-2000. There is interest in reconsidering these outcomes a decade later, during which statistics and probability has become a recognised strand of the Australian mathematics curriculum. Using a new data-set of over 7000 student responses from middle-years students in different parts of Australia during the period 2007-2009, the nature of the hierarchy was confirmed. Longitudinal analysis identified how students performed across time against the hierarchy. Suggestions are made for systems and teachers about realistic expectations for middle-years students, and possible curriculum challenges.

Keywords: *Statistical literacy; Rasch modelling; Middle years; Longitudinal analysis.*

1. INTRODUCTION

The place of statistical literacy within the field of statistics education has been a matter of debate and redefinition for at least 20 years. In fact the antecedents of statistical literacy have roots 30 years ago with a collaboration of the American Statistical Association (ASA) and the National Council of Teachers of Mathematics (NCTM) that produced the *Quantitative Literacy Series*, including titles *Exploring Data* (Landwehr & Watkins, 1986), *Exploring Probability* (Newman, Obremski, & Scheaffer, 1987), *The Art and Techniques of Simulation* (Gnanadesikan, Scheaffer, & Swift, 1987), *Exploring Surveys and Information from Samples* (Landwehr, Swift, & Watkins, 1987), and *Exploring Measurements* (Barbella, Kepner, & Scheaffer, 1994). Steen (2001, p. 5) claimed that the name for the series was chosen because at the time the word “statistical” was considered too threatening for the intended audience.

With the growing acceptance of data and chance as a part of national school curricula, beginning with the NCTM (1989), statistical literacy began to be recognized as important, especially by those in the emerging field of statistics education. Among the significant issues, however, was the distinction among statistical reasoning, statistical thinking, and statistical literacy: three phrases that have been the basis for Statistical

Reasoning, Thinking and Literacy (SRTL) Forums every two years since 1999 <srtil.tos.auckland.ac.nz>. Statistical literacy was also one of the three foci of Ben-Zvi and Garfield's (2004) opening chapter of their book on the subject, which suggested the following description of statistical literacy as the lowest of the three levels of mental involvement with statistics.

Statistical literacy includes basic and important skills that may be used in understanding statistical information or research results. These skills include being able to organize data, construct and display tables, and work with different representations of data. Statistical literacy also includes an understanding of concepts, vocabulary, and symbols, and includes an understanding of probability as a measure of uncertainty. (p. 7)

In the same book, however, Gal's (2004) description of statistical literacy immediately expanded on the above to refer broadly to two interrelated components, primarily

- (a) people's ability to interpret and critically evaluate statistical information, data-related arguments, or stochastic phenomena, which they may encounter in diverse contexts, and when relevant
- (b) their ability 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. (p. 49)

This description includes aspects of statistical reasoning, where "[r]easoning means understanding and being able to explain statistical processes and being able to fully interpret statistical results", and aspects of statistical thinking, where "statistical thinkers are able to critique and evaluate results of a problem solved or a statistical study" (Ben-Zvi & Garfield, 2004, p. 7).

Across the research field, four populations of interest have emerged in relation to the need for statistical literacy: university students; the general adult population; teachers at the school level; and school students. Statistics education obtained at university covers a wide range of needs, from specialist statisticians, to people who routinely use or encounter statistics in a professional capacity, such as researchers in health or psychology, policy makers or marketing specialists. Zieffler, Garfield, Alt, Dupuis, Holleque, and Chang (2008) provided a useful review of the existing research about the teaching and learning of introductory statistics at the college level and concluded that statistics remained for many a difficult subject and that faulty statistical reasoning was pervasive even after taking statistics courses. Tishkovskaya and Lancaster (2012) further reviewed the challenges, innovations and reforms at this level, whereas other studies looked more specifically at team-based learning (St. Clair & Chihara, 2012), at units based entirely on contexts from the media (Budgett & Pfannkuch, 2010), and at student attitudes and performance in web-based, fully online, and flipped teaching conditions (Gundlach, Richards, Nelson, & Levesque-Bristol, 2015). Brown, David, Moltchanova, Seddon, and Harlow (2014) described a program of consolidating help with statistical literacy across campuses and across subject domains at their New Zealand University. Schield (2002) suggested three kinds of statistical literacy he wanted for tertiary students: chance-based, fallacy-based, and correlation-based. Later he focussed on context, assembly (from choice of method), randomness, and error (or bias) (Schield, 2010).

Gal's (2004) concern was about adults in society, as was that of Wallman (1993), Moreno (2002), Utts (2003), and Scheaffer (2003). There is broad agreement that in an increasingly quantitative world having an understanding of statistical information and being able to interpret and act on that information is essential (United Nations Economic

Commission for Europe, 2012). Petty (2010, 2014) described her novel approach of using social media to get the message of statistical literacy across to both students and the wider community. If it is important for adult citizens to be statistically literate, then what outcomes from school education should provide the necessary foundation?

This question leads to a consideration of the third population of interest: teachers. There are two aspects of teachers' statistical literacy that are important. First, teachers' influence in the classroom involves the pedagogical content knowledge (PCK) (Shulman, 1987) required to engage students in the statistical literacy aspects of the mathematics curriculum (e.g., Watson & Nathan, 2010a), in particular being able to extend concepts, such as sampling, to controversial media contexts (e.g., Watson & Nathan, 2010b). Vita, Santana, Cazorla, and Oliveira (2010) used Activity Theory to analyse 34 teachers' actions in an activity based on the "Vitruvian Man", whereas Aizikovitsh-Udi and Kuntze (2014) presented two case studies analysing the impact of critical thinking on experienced secondary teachers' statistical literacy. Hay (2010) analysed transcripts of teacher interviews to produce a concept map of teachers' perceptions of best practice in relation to statistical literacy. It has been recognised that having higher levels of statistical PCK in teachers leads to better student learning outcomes (Callingham, Carmichael, & Watson, 2016). Second in relation to the population of teachers, interest has recently developed in teachers' understanding of statistical literacy in a context where it is necessary to interpret data particularly from student testing programs but also other sources (Chick, Pierce, & Wander, 2014; Pierce, Chick, Watson, Les, & Dalton, 2014). This could also be classified as workplace statistical literacy rather than the knowledge required to teach statistical literacy to students.

The final population of interest is that of school students. Many of the early suggestions for the school level looked more broadly at statistics learning (e.g., Boland, 2002; Scheaffer, Watkins, & Landwehr, 1998) rather than statistical literacy itself. Watson (2006) framed her research on the following definition of statistical literacy for all school students.

Statistical literacy is the meeting point of the data and chance curriculum and the everyday world, where encounters involve unrehearsed contexts and spontaneous decision-making based on the ability to apply statistical tools, general contextual knowledge, and critical literacy skills. (p. 11)

Watson's follow-up research and suggestions for learning to develop statistical literacy drew from the work of Holmes (1980) and Friel and Bright (1998): sampling, graphs, average, chance and beginning inference, with special attention paid to variation. More recently participation in specific classroom investigations involving statistical questions has been advocated for building statistical literacy. For example, Leavy and Hourigan (2016) worked with Grade 6 students in two "mystery" scenarios requiring creation of data distributions, English and Watson (2015) considered question posing as part of statistical literacy with Grade 4 students, and Marshall and Swan (2006) used M&Ms with younger children. Sproesser, Kuntze, and Engel (2014) considered various factors impacting on statistical literacy competence, including grades in mathematics, verbal and non-verbal cognitive ability, socio-economic status, conceptual knowledge of probability, and conceptual knowledge of functional reasoning. All factors showed significant relationships for students in Grades 8 or 9. This school level research led to a consideration of how best to assess statistical literacy.

2. ASSESSMENT OF STATISTICAL LITERACY

Following the discussion on the content of statistical literacy and what aspects should be taught and learned at different levels of education, the issue of assessing that learning arises. Garfield and Gal (1997) began the exchange of ideas in the wider field of statistics education with a diverse range of suggestions. Although most suggestions addressed the tertiary level and material from formal statistics courses, there were some links to what has become statistical literacy. The general discussion of assessment of statistical understanding continued at the tertiary level in the book edited by Bidgood, Hunt, and Jolliffe (2010), which included specific examples for statistical literacy by Schield (2010) and Budgett and Pfannkuch (2010). Their assessment techniques for classroom settings included multiple choice questions, exercises comparing statistics in graphs and tables, assignments based on judging statistical studies and associated media reports, construction of statistical reports, and more formal university examinations. Bidgood (2014) was especially supportive of using authentic “real world” contexts and data for assessment. Kaplan and Thorpe (2010) took this further in a pilot project assessing adults’ statistical literacy beyond the classroom. At the school level Watson (1997) considered assessment tasks based on questionable media articles and suggested hierarchical levels of performance. A similar approach was taken with pre-service teachers (Watson, 2011). Using complex tasks in rich contexts is appropriate and necessary for teachers but is considerably more difficult when applied to large-scale assessment.

2.1. INSTRUMENTS FOR LARGE-SCALE ASSESSMENT

Green (1982) was one of the first in the field to explore the understanding of a large number of school students with a survey. His work in probability was a landmark and many of his items for school students have continued to be used across the years. Fischbein and Gazit (1984) surveyed students across Grades 5 to 7, also providing probability items for future researchers. At the tertiary level, the work of Konold and Garfield (1992) on intuitive statistical reasoning laid the foundation for many of the later developments, e.g., the ARTIST project (Garfield, delMas, & Chance, 2002), the CAOS project (delMas, Garfield, Ooms, & Chance, 2007), and the GOALS project (Garfield, delMas, & Zieffler, 2010), also in a large scale context. Arguing that these instruments required updating to reflect the importance of simulation to statistical literacy, Ziegler (2014) described the development of the Basic Literacy in Statistics (BLIS) assessment instrument. A feature was the inclusion of questions based on the presentation of the results of simulation.

The work of Watson and Callingham (2003, 2005) and Callingham and Watson (2005) moved the statistical literacy focus to the school level, employing some items adapted from earlier tertiary assessments (e.g., Garfield, 2003; Konold & Garfield, 1992; Tversky & Kahneman, 1983), as well as items used by others with school students (e.g., Fischbein & Gazit, 1984; Green, 1982; Jacobs, 1999). Keeping in mind the importance of context and critical thinking, items were also developed based on questionable reports from the media (e.g., Watson, 1997, 1998). All items used in the earlier studies, their rubrics, and details of their origins are included in Watson and Callingham (2003, 2005).

2.2. A STATISTICAL LITERACY CONSTRUCT

The analyses of Watson and Callingham (2003, 2005) suggested a single underlying variable for statistical literacy, which could be described in six levels. These are

summarised in Table 1. Increasing levels reflect more knowledge of statistical procedures, for example related to tables, graphs, averages and probabilities; greater appreciation of variation and context as encountered in tasks; and increased critical thinking in decision making acknowledging uncertainty. Only at the highest level are students able to apply proportional reasoning understanding to complete statistical tasks.

Table 1. Statistical literacy hierarchy

Level	Name	Characteristics of Level
6	Critical Mathematical	Critical, questioning engagement with context, using proportional reasoning particularly in media or chance contexts, showing appreciation of the need for uncertainty in making predictions, and interpreting subtle aspects of language.
5	Critical	Critical, questioning engagement in familiar and unfamiliar contexts that do not involve proportional reasoning, but which do involve appropriate use of terminology, qualitative interpretation of chance, and appreciation of variation.
4	Consistent Non-critical	Appropriate but non-critical engagement with context, multiple aspects of terminology usage, appreciation of variation in chance settings only, and statistical skills associated with the mean, simple probabilities, and graph characteristics.
3	In-consistent	Selective engagement with context, often in supportive formats, appropriate recognition of conclusions but without justification, and qualitative rather than quantitative use of statistical ideas.
2	Informal	Only colloquial or informal engagement with context often reflecting intuitive non-statistical beliefs, single elements of complex terminology and settings, and basic one-step straightforward table, graph, and chance calculations.
1	Idiosyncratic	Idiosyncratic engagement with context, tautological use of terminology, and basic mathematical skills associated with one-to-one counting and reading cell values in tables.

The construct described in Table 1 was identified based on responses from over 3,000 students in Years (Grades) 3 to 9 (Watson & Callingham, 2003) using archived data collected between 1993 and 2000. Using 23 of the original items and 26 new items, new surveys of 673 students in years 5 to 10 confirmed the original hierarchical structure (Callingham & Watson, 2005; Watson & Callingham, 2005). In that study the mean abilities across the years were shown generally to improve across the years of schooling. The largest proportion of responses at Year 5 was at the Inconsistent level (Level 3), whereas from Year 6 onwards the largest fraction of responses was at the Consistent Non-Critical level (Level 4), suggesting that movement into the higher Critical levels was more difficult and required focused and targeted intervention. These upper levels also require greater depth of statistical reasoning and thinking, including drawing inferences and justifying these statistically.

Over the decade since these analyses were carried out, the profile of statistics and probability has increased in Australia, as well as in other countries around the world. In Australia, for example, the name of that part of the mathematics curriculum was changed from “Chance and Data” to “Statistics and Probability” (Australian Curriculum, Assessment and Reporting Authority [ACARA], 2015) and in New Zealand, Statistical Literacy was included as a specific sub-section of the *Mathematics and Statistics* curriculum (Ministry of Education, 2009). The question arises as to whether generally students have had greater exposure to statistical concepts, resulting in higher levels of

performance across grades. The StatSmart project (Callingham & Watson, 2007) provided the opportunity to reconsider the construct of statistical literacy in an environment where the curriculum had changed with a stronger focus on statistics.

Against that background, the research questions addressed in the current study are:

(1) To what extent does the analysis of the data from the current study support the uni-dimensional statistical literacy hierarchy identified in earlier studies (Callingham & Watson, 2005; Watson & Callingham (2003, 2005)?

(2) To what extent does students' understanding of statistical literacy change as they move through school?

3. METHODOLOGY

The StatSmart project (Callingham & Watson, 2007) aimed to improve statistical learning outcomes for students in the middle years through targeted professional learning of teachers. Over the three years of the study, from 2007 to 2009, over 4000 students in Years 5 to 11 across three Australian states were assessed using three different test forms that included many of the items used by the authors in their earlier studies. In most cases students were surveyed longitudinally three times: before a statistics unit taught by a StatSmart teacher, after the unit but in the same year as being taught by the StatSmart teacher, and one year later if still in the school, regardless of teacher in the meantime. For over 200 of the students, who were in the project a year after their third survey, it was possible to administer a fourth or fifth survey and trial some new items in surveys in conjunction with items linked to the other survey forms used in the project. Altogether there were five different survey forms that included a total of 71 items, all addressing aspects of statistical literacy. The underlying construct was theorised as being the statistical literacy hierarchy identified by Watson and Callingham (2003, 2005) and Callingham and Watson (2005).

The tests were administered in students' classrooms by their class teacher. A rotating design was used so that across the three key tests (pre-test, post-test, longitudinal follow up) no student saw the same test twice. Tests were undertaken in the first half of the school year (April/May in Australia), and at the end of the school year (Late November). Overall there were six test periods. The April/May tests were given only to students new to the project because they were in a class taught by a project teacher. Tests at the end of the school year were given to these same students as their post-test, and to as many students as could be identified from the previous year as a longitudinal follow up. The full rotating test design is described in Callingham and Watson (2007). Because the teachers involved in the study changed their classes and teaching year groups from year to year some students who had completed the three key tests were in a class taught by a StatSmart teacher by chance. To accommodate teachers' requests for a test form that they could use with students who had completed all three key tests, two further test forms were developed, providing the opportunity to trial new items. Completed tests were sent to the research team, and coded by trained research assistants using rubrics developed by the research team. The researchers additionally marked a sample of tests to ensure coding consistency.

In addition to many items that were identical or very similar to items used in previous studies, some new items had been developed during the StatSmart project. In particular there were new items that focussed mainly on probabilistic ideas in various social contexts, such as a lottery (RUTH, TATS), dependence or independence of factors related to diseases (T2X2, SKINR, STOMR), births in a hospital (HOSP), tossing coins (TED), a family leaving a neighbourhood (F2.3), and a newspaper article about getting out of bed

on the left side being lucky (BED, JMES, FENG, PSYC). These items were based on questions used in interview protocols or were items adapted from other survey research not included in Watson and Callingham (2003). Each item was coded using a rubric that described the increasing sophistication of the response. The items and their rubrics are included in the Appendix, with references to their original sources.

All five tests included five common items, and another set of five items linked the three initial test forms that provided the pre-, post- and follow-up tests. These link items allowed the use of Rasch analysis approaches to create a common scale. This approach to analysis places both test takers (persons) and items on the same metric, measured in logits, the logarithm of the odds of success. The person values are termed ability measures, and the item values are difficulty measures.

Rasch analysis provides evidence of construct validity through examination of the fit of items to the underlying model—items that fit the underlying model can be considered as measuring a single underlying construct (Bond & Fox, 2015) that can be interpreted against the theorised variable. Hence the nature of the scale could provide evidence of whether the construct of statistical literacy had changed over the 10-year period since it was identified.

Responses to every item at every test event were combined into a single dataset resulting in 7057 data cases. For this paper, students in Years (grades) 11 and 12 were removed from the data set because of their low numbers. The resulting data set of 7024 responses was analysed using the Partial Credit Model (Masters, 1982) with Winsteps 3.92 software (Linacre, 2016). Output was examined for fit to the model using the infit mean square value, which has an ideal value of 1 and is generally acceptable if it lies between 0.5 and 1.5 (Linacre, 2016). Using an identical process to the original study (see Callingham & Watson, 2005, for details) the variable obtained through this process was segmented into bands or levels. The items within each identified band were then considered in comparison to those items identified in the equivalent level of the statistical literacy hierarchy in the earlier studies to re-validate the nature of the variable.

Once the scale was established, the person ability measures were obtained for each of the 7024 data cases. These values were exported into SPSS for further analysis. In particular, there was interest in examining the proportions of persons at each level of the variable. Because the same students undertook at least three separate tests, this provided the opportunity to examine change over time, or a genuine developmental measure. In a set of separate Rasch analyses, for each of the six rounds of testing that took place over three years, person ability measures were obtained, anchored to the original test so that all values were on the same scale. These were also exported into SPSS, matched by individual student and combined into a large data file in which individual students could be tracked across time.

4. RESULTS

4.1. STATISTICAL LITERACY CONSTRUCT

In terms of fit to the Rasch model, the combined set of 71 items showed good fit of both infit mean square (IMNSQ = 0.99, zStd = -0.5) and outfit (OMNSQ = 0.98, zStd = -0.6) indicating that the items worked together consistently to provide a sound measurement scale. One item, TRV3, showed some misfit that could potentially degrade the measurement qualities of the scale (IMNSQ = 1.66, Infit zStd = 9.9; OMNSQ = 1.94, Outfit zStd = 9.9). Several other items had unstandardised fit values of close to 1.0 but large standardised values. As discussed by Linacre (2016) high z values can occur when

sample size is large, as in this instance. Linacre indicated that whereas the Infit and Outfit values provide information about the utility of the fit to the Rasch model, the standardized values, *z* scores, provide information about the precision of the fit. When large samples are involved, the ideals of the Rasch model are less likely to be met because of the imprecision of the measurement process. For the intended purpose of the study reported here, the fit to the model of items other than TRV3 was satisfactory. In terms of person fit, the overall fit to the model of person responses was good (IMNSQ = 1.06, Infit *z*Std = 0.1; OMNSQ = 0.99, Outfit *z*Std = 0.0) suggesting that overall the people attempting these items did so consistently.

The Rasch person reliability measure was high at 0.87, with a person separation index of 2.56. The low separation index has implications for the interpretation of the information, and will be discussed further in the Discussion section. The Rasch item reliability was very high at 1.00, with a large separation index of 19.31. These statistics imply that the item hierarchy can be considered as sound, with items covering a wide range of difficulty. In summary, the items and persons appeared to provide useful information about the underlying theorised construct of Statistical Literacy. To examine the nature of that construct, the variable map, called a Wright map (Wilson & Draney, 2000), produced by the software, was considered (see Figure 1).

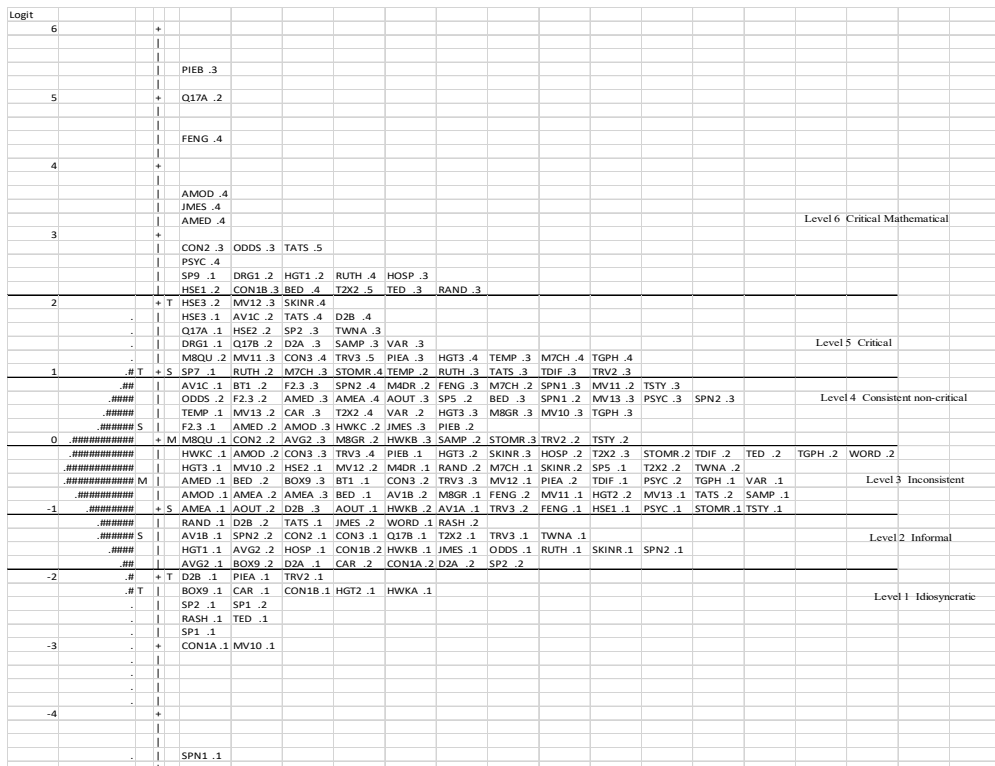


Figure 1. Wright map of all statistical literacy items

Each # represents 78 responses and each . represents 1 to 77 responses

The scale on the far left is in logits, the units of Rasch measurement. The # symbols represent the distribution of persons, with each # standing for 13 people. The items are shown on the right of the graph, with the decimal number showing the code associated with the item rubric. Hence FENG.4 shows the position on the hierarchy of the code 4

level of response. The items are distributed along the variable with the easiest codes to achieve at the bottom of the map and the hardest codes to achieve at the top. Similarly the persons are distributed from the lowest performances at the bottom (low ability) to the highest at the top.

In order to compare the nature of the variable with the earlier construct of statistical literacy the map was examined for places where there seemed to be a change in the difficulty level and the items had common cognitive demands. The initial tentative division of the variable was then compared with the previous construct (Callingham & Watson, 2005) to confirm the placement of the levels. In general there had been little change in the relative placement of the common items over the 10 years since the initial analysis, indicating that the variable was stable. Finally, the cognitive demands of the item codes at each level were considered to see whether the descriptors were still valid, especially taking into consideration the new items. In this way the variable was segmented into six levels as shown in Figure 1, labelled according to the levels of the original hierarchy. Overall, this qualitative analysis of the variable confirmed the nature of the statistical literacy hierarchy, suggesting that the items measured a stable, enduring trait in a consistent fashion.

4.2. DISTRIBUTION OF STUDENTS ACROSS LEVELS

Table 2 shows the composition of the responses by year and gender. The number of responses in each year (grade) level was sufficient to use to compare the distribution of students across levels by year, allowing for the development of tentative benchmarks.

Table 2. Composition of the student sample (N = 7024)

Year	Male (n)	% male in year	Female (n)	% female in year	Total (n)	% of sample
5	92	44.2	116	55.8	208	3.0%
6	535	53.9	457	46.1	992	14.1%
7	925	51.7	863	48.3	1788	25.5%
8	1075	49.9	1079	50.1	2154	30.7%
9	565	53.1	499	56.9	1054	15.0%
10	341	41.2	487	58.8	828	11.8%

Having established that the variable appeared to be stable across time, there was interest in examining the distribution of students across the levels. The measures of student ability were grouped by year and cut scores for persons were established at the same values in logits as the placement of the lines on the variable map. The proportions of students in every year at each level of the hierarchy were determined. Table 3 shows the percentage of students at each level grouped by year. The heavier shading shows the highest percent for each year group, and the lighter shading shows the next highest percent.

Table 3. Percent of students in levels of the hierarchy by year (grade) group

Year	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6
Yr 5	19.7%	21.2%	41.8%	17.3%	0.0%	0.0%
Yr 6	17.4%	24.1%	43.6%	14.7%	0.2%	0.0%
Yr 7	8.2%	14.2%	38.7%	36.4%	2.5%	0.0%
Yr 8	7.2%	11.0%	32.3%	41.6%	7.6%	0.4%
Yr 9	4.3%	7.5%	24.9%	47.9%	14.6%	0.8%
Yr 10	4.4%	6.9%	28.0%	46.4%	13.2%	1.2%

In Years 5 and 6, most students are located in the Inconsistent level (Level 3), but from Year 8 onwards, the Consistent Non-critical level (Level 4) becomes more, with almost the same proportion of Year 7 students in Level 3 and Level 4. Although the analysis cannot be exactly compared with the original data, in general the patterns are similar. Students seem to move up to Level 4, Consistent Non-critical, but find it difficult to reach the higher levels. The implications of this finding are explored in the Discussion.

In addition to the levels grouping, there was interest in considering the relative measure of performance, the ability measure, of each year group. This analysis provides an indication of how performance against the variable might change across the years of schooling. Figure 2 shows the mean ability measures ± 2 standard errors by year. Of particular interest is that there does not appear to be the commonly observed plateau in performance as students move from primary school in Year 6 to secondary school in Year 7. Rather there appears to be a plateau from Year 9 to 10, and a one-way ANOVA using a Bonferroni posthoc test indicated significant differences between groups ($F = 130.027$, $df = 7$, $p < 0.01$) with only the differences from Year 5 to Year 6 and Year 9 to Year 10 showing no statistical significance. This finding is also considered further in the discussion section.

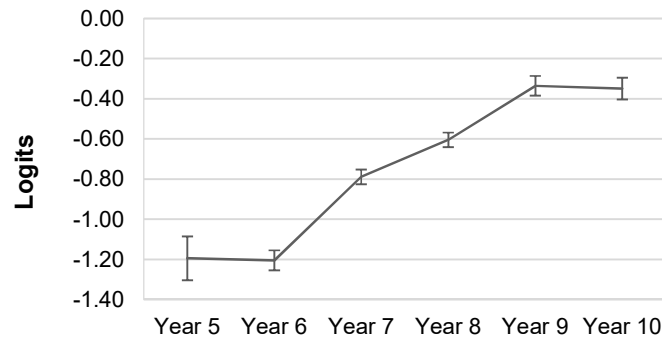


Figure 2. Mean ability measure by year level with associated ± 2 standard errors

4.3. LONGITUDINAL ANALYSIS

The outcomes for individual students across time were computed and organised into a single data file. There were two phases of data available, which each had three test events. Tests 1 and 2 were undertaken in the same year, approximately 6 months apart, before and after the students had been taught a statistics unit as part of their normal mathematics curriculum with a StatSmart teacher. They were tested again approximately 1 year after Test 2, regardless of whether they had continued to be taught by a StatSmart teacher. Using the same logit values as the previous analysis, proportions of students by year and level of the statistical literacy hierarchy were calculated. The results of this analysis are shown in Figure 3. It should be noted that the year level is the year at the start of the StatSmart testing process, and that by Test 3 students were in the next year, that is, students in Year 6 at Tests 1 and 2 were in Year 7 for Test 3.

The impact of the StatSmart project can be seen by considering those reaching Level 4, consistent Non-critical. At Test 2 across all year levels, there was a drop in the proportion of students in Levels 1 to 3, and more students reaching Level 4. At Test 3, however, the patterns are mixed. Students who started the project in Year 5 appeared to have lost some gains by Test 3, when they were in Year 6. A similar pattern is shown by

students who started in Year 9 and were in Year 10 by Test 3. Students who started in Year 6 and Year 8 appeared to maintain their improvement in Test 3, and students who started in Year 7 at least maintained the gains made by Test 2. Remembering that there was no special intervention between Test 2 and Test 3, these differences may have been due to the curriculum taught, and the backgrounds of the teachers who took the relevant classes. It is not possible to tease out all of the potential variables that might have impacted on these students as they progressed through school, but overall the StatSmart intervention through teachers' professional development did appear to support students' development of statistical literacy, particularly in the middle years of schooling.

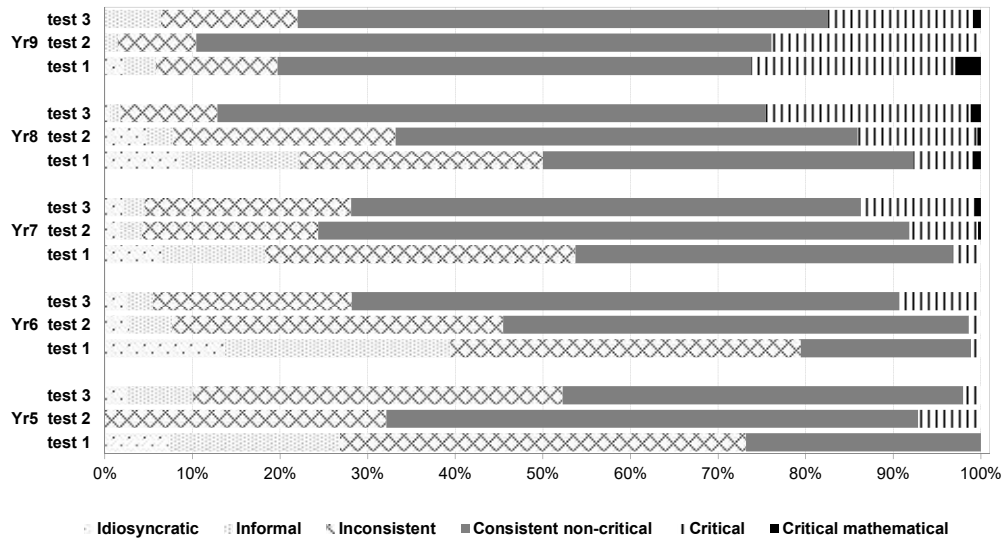


Figure 3. Proportions in levels of matched students over time by year group at project start

5. DISCUSSION

The purpose of this study was to re-examine the statistical literacy hierarchy established 10 years prior to the StatSmart study (Callingham & Watson, 2005; Watson & Callingham, 2003, 2005). Although the comparison with the earlier hierarchy was, of necessity, qualitative, it appears that the construct had remained very stable across time. The descriptors of demands at the different levels still appeared appropriate, including new and revised items. This finding suggests that the hierarchy provides a sound basis for curriculum development at the school level. In addition, the longitudinal analysis indicates that the statistical literacy hierarchy can be used to monitor progress, and that students can improve their performance with targeted and appropriate teaching.

There are some concerns, however. It appears that students tend to plateau at Level 4, Consistent Non-Critical, of the hierarchy. At this level students can undertake the basic mathematical calculations of central tendency, for example, recognise the difference between a sample and a population, and variation when throwing a die or using other chance devices. They do not, however, use these skills to make any critique of statistics presented to them or to interpret statistics in familiar or unfamiliar contexts. In this respect little seems to have changed during the 10 years between the various studies.

Although somewhat disappointing, there may be good underlying reasons for this plateau. Cognitive developmental theorists starting with Piaget (e.g., Inhelder & Piaget,

1958) point to the later years of schooling as the period people begin to deal with abstractions and generalisations, suggesting that true critical thinking is unlikely to occur in the middle years of schooling. It is in these years, however, that the foundations are laid for adult thinking, and the findings in relation to statistical literacy suggest that schools could do more to familiarize students with the necessary vocabulary and mathematical skills on which they will draw as they develop further. When considered in the broader context of statistical literacy education, these Level 4 skills seem to be similar to those required by Ben-Zvi and Garfield (2004) quoted in the introduction. It does not, however, meet the expectations of Gal (2004) for adults, also quoted earlier, which require a much more critical stance. The capacity to use statistical knowledge to make critical judgments about a situation, taking account of the context and the statistical approaches does not appear to be emphasised within the Australian school system.

Of concern also is the lack of measured growth from Year 5/6 to Year 9/10, improving only slightly from -1.20 logits to -0.4 logits. The “clumping” of students in the lower half of the scale shown in Figure 1, and the mean person measure being below the mean item measure (0 logits), suggests that the students found the tests they took relatively difficult. Visually comparing the distribution of students with those who undertook the initial assessments (Callingham & Watson, 2005) suggests that fewer students in this study were reaching the upper levels of the hierarchy. Unfortunately, for technical reasons no true anchored analysis can be undertaken so this conjecture must remain open. The lack of growth does, however, suggest that Australian schools could be doing more to develop students’ statistical understanding. In addition, it might be worth considering developing a more targeted instrument to illuminate statistical understanding in more detail in the lower middle years.

Of further interest is the relative lack of growth from Year 5 to Year 6, and from Year 9 to Year 10, but continued growth, albeit small, across the transition from primary to secondary school (Year 6 to Year 7). This pattern is unexpected, because there is usually a drop in performance at the transition to secondary school. It may be that only limited statistics was taught in the primary years, despite these students being part of a project focused on statistics teaching. Certainly anecdotal evidence suggested that teachers who taught in the lower years of secondary schools were most enthusiastic and tried different approaches to teaching statistics.

The plateau from Year 9 to Year 10, however, suggests that the skills developed by Year 9 are not further developed in Year 10. Another reason could be that as they move into secondary school, students are exposed to greater amounts of statistical information across the curriculum. This information includes data from science experiments, geography, history, health science, and so on. Much of this information is presented as frequency counts or straightforward percentages of a quantity, such as in a pie chart. This exposure is likely to help the development of general statistical skills. What appears to be missing, however, is a sense of critique of information, particularly in a statistical context. If students are to move to higher levels of the statistical literacy hierarchy then the skills of reading data, drawing on proportional reasoning, need to be deliberately developed in a focused and targeted manner. This kind of responsibility would appear to extend beyond the mathematics classroom, because it is as important in other areas of the curriculum as in mathematics. Strategies such as using media articles (Watson, 2006), investigative tasks (Leavy & Hourigan, 2016; Marshall & Swan, 2006), and greater use of real world contexts (Bidgood, 2014), are needed.

Although these results should be considered tentatively, the longitudinal analysis indicates that the skills of statistical literacy can be developed, and that when students are taught in a deliberate and focused way they can retain much of their new knowledge, with

limited additional intervention. At the upper years of the middle years, Years 9 and 10, particular attention should be paid to statistical reasoning and thinking in order to provide a platform for developing the higher levels of the hierarchy.

It is somewhat disappointing that despite the work done in the intervening 10 years between the first identification of the statistical literacy hierarchy and the StatSmart project, little appears to have changed in terms of students' overall statistical development. Clearly there is still much more focus needed on statistics education at the school level if the populace is to become genuinely statistically literate.

ACKNOWLEDGEMENT

This research was funded by Australian Research Council grant number LP0669106.

REFERENCES

- Aizikovitsh-Udi, E. & Kuntze, S. (2014). Critical thinking as an impact factor on statistical literacy – Theoretical frameworks and results from an interview study. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Proceedings of the Ninth International Conference on Teaching Statistics: Sustainability in Statistics Education*. Voorburg, The Netherlands: International Association for Statistical Education.
[Online: iase-web.org/Conference_Proceedings.php?p=ICOTS_9_2014]
- Australian Curriculum Assessment and Reporting Authority (2015). *Australian Curriculum: Mathematics, Version 7.4, 30 March 2015*. Sydney, NSW: ACARA.
- Barbella, P., Kepner, J., & Scheaffer, R. L. (1994). *Exploring measurements*. [Quantitative Literacy Series] Palo Alto, CA: Dale Seymour.
- Batanero, C., Estepa, A., Godino, J. D., & Green, D. R. (1996). Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education*, 27, 151–169.
- Ben-Zvi, D. & Garfield, J. (2004). Statistical literacy, reasoning, and thinking: Goals, definitions, and challenges. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 3–25). Dordrecht, The Netherlands: Kluwer.
- Bidgood, P. (2014). Towards statistical literacy – Relating assessment to the real world. In K. Makar, B. deSousa, & R. Gould (Eds.), *Proceedings of the Ninth International Conference on Teaching Statistics: Sustainability in Statistics Education*. Voorburg, The Netherlands: International Statistical Institute.
[Online: iase-web.org/Conference_Proceedings.php?p=ICOTS_9_2014]
- Bidgood, P., Hunt, N., & Jolliffe, F. (2010). *Assessment methods in statistical education: An international perspective*. Chichester, UK: Wiley.
- Boland, P. J. (2002). Promoting statistics thinking amongst secondary school students in the national context. In B. Phillips (Ed.), *Proceedings of the 6th International Conference on Teaching of Statistics: Developing a Statistically Literate Society*. Voorburg, The Netherlands: International Statistical Institute.
[Online: iase-web.org/Conference_Proceedings.php?p=ICOTS_6_2002]
- Bond, T. G. & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). New York: Routledge.
- Brown, J., David, I., Moltchanova, E., Seddon, H., & Harlow, J. (2014). Improving statistical literacy at university. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Proceedings of the Ninth International Conference on Teaching Statistics: Sustainability in Statistics Education*. Voorburg, The Netherlands: International Association for Statistical

- Education and the International Statistical Institute.
[Online: iase-web.org/Conference_Proceedings.php?p=ICOTS_9_2014]
- Budgett, S. & Pfannkuch, M. (2010). Assessing students' statistical literacy. In P. Bidgood, N. Hunt, & F. Jolliffe (Eds.), *Assessment methods in statistical education: An international perspective* (pp. 103–121). Chichester, UK: Wiley.
- Callingham, R., Carmichael, C., & Watson, J. (2016). Explaining student achievement: The influence of teachers' pedagogical content knowledge in statistics. *International Journal of Science and Mathematics Education*, 14(7), 1339–1357.
[Online: doi.org/10.1007/s10763-015-9653-2]
- Callingham, R. & Watson, J. M. (2005). Measuring statistical literacy. *Journal of Applied Measurement*, 6(1), 19–47.
- Callingham, R. & Watson J. M. (2007). Overcoming research design issues using Rasch measurement: The StatSmart project. In P. Jeffery (Ed.), *Proceedings of the AARE annual conference, Fremantle*.
[Online: www.aare.edu.au/publications-database.php/5329/overcoming-research-design-issues-using-rasch-measurement-the-statsmart-project]
- Chick, H. L., Pierce, R., & Wander, R. (2014). Sufficiently assessing teachers' statistical literacy. In K. Makar, B. deSousa, & R. Gould (Eds.), *Proceedings of the Ninth International Conference on Teaching Statistics: Sustainability in Statistics Education*. Voorburg, The Netherlands: International Statistical Institute.
[Online: iase-web.org/icots/9/proceedings/pdfs/ICOTS9_7C3_WANDER.pdf]
- delMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal*, 6(2), 28–58. [Online: iase-web.org/Publications.php?p=SERJ_issues]
- English, L. D. & Watson, J. M. (2015). Statistical literacy in the elementary school: Opportunities for problem posing. In F. Singer, N. Ellerton, & J. Cai (Eds.), *Problem posing: From research to effective practice* (pp. 241–256). Dordrecht: Springer.
- Fischbein, E. & Gazit, A. (1984). Does the teaching of probability improve probabilistic intuitions? An exploratory research study. *Educational Studies in Mathematics*, 15, 1–24.
- Friel, S. N. & Bright, G. W. (1998). Teach-Stat: A model for professional development in data analysis and statistics for teachers K–6. In S.P. Lajoie (Ed.), *Reflections on statistics: Learning, teaching, and assessment in grades K-12* (pp. 89–117). Mahwah, NJ: Lawrence Erlbaum.
- Gal, I. (2004). Adults' statistical literacy: Meanings, components, responsibilities. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 47–78). Dordrecht, The Netherlands: Kluwer.
- Garfield, J. B. (2003). Assessing statistical reasoning. *Statistics Education Research Journal*, 2(1), 23–38. [Online: [iase-web.org/documents/SERJ/SERJ2\(1\).pdf](http://iase-web.org/documents/SERJ/SERJ2(1).pdf)]
- Garfield, J., delMas, R., & Chance, B. (2002). Assessment resource tools for improving statistical thinking (ARTIST). [Online: apps3.cehd.umn.edu/artist/index.html]
- Garfield, J., delMas, R., & Zieffler, A. (2012). Developing statistical modelers and thinking in an introductory, tertiary-level statistics course. *ZDM – The International Journal on Mathematics Education*, 44(7), 883–898.
[Online: doi.org/10.1007/s11858-012-0447-5]
- Garfield, J. & Gal, I. (1997). *The assessment challenge in statistics education*. Amsterdam: IOS Press and The International Statistical Institute.
- Gnanadeskin, M., Scheaffer, R. L., & Swift, J. (1987). *The art and techniques of simulation*. [Quantitative Literacy Series] Palo Alto, CA: Dale Seymour.
- Green, D. R. (1982). *Probability concepts in 11-16 year old pupils*. Loughborough, UK: Center for Advancement of Mathematical Education in Technology.

- Gundlach, E., Richards, K. A. R., Nelson, D., & Levesque-Bristol, C. (2015). A comparison of student attitudes, statistical reasoning, performance, and perceptions for web-augmented traditional, fully online, and flipped sections of a statistical literacy class. *Journal of Statistics Education* 23(1).
[Online: ww2.amstat.org/publications/jse/v23n1/gundlach.pdf]
- Hay, I. (2010). Teachers' perceptions of best practice in statistical literacy education. In C. Reading (Ed.), *Proceedings of the 8th International Conference on Teaching Statistics: Data and Context in Statistics Education: Towards an Evidence-Based Society*. Voorburg, The Netherlands: International Association for Statistical Education.
[Online: iase-web.org/documents/papers/icots8/ICOTS8_C158_HAY.pdf]
- Holmes, P. (1980). *Teaching Statistics 11-16*. Slough, UK: Schools Council Publications and W. Foulsham & Co.
- Inhelder, B. & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. London: Routledge and Kegan Paul.
- Jacobs, V. R. (1999). How do students think about statistical sampling before instruction? *Mathematics in the Middle School*, 5(4), 240–263.
- Kahneman, D., & Tversky, A. (1972). Subjective Probability: A judgment of representativeness. *Cognitive Psychology*, 3, 430–454.
- Kaplan, J. J. & Thorpe, J., (2010). Post secondary and adult statistical literacy: Assessing beyond the classroom. In C. Reading (Ed.), *Proceedings of the 8th International Conference on Teaching Statistics: Data and Context in Statistics Education: Towards an Evidence-Based Society*. Voorburg, The Netherlands: International Association for Statistical Education.
[Online: iase-web.org/documents/papers/icots8/ICOTS8_5E3_KAPLAN.pdf]
- Konold, C. & Garfield, J. (1992). *Statistical reasoning assessment: Part 1. Intuitive thinking*. Amherst, MA: Scientific Reasoning Research Institute, University of Massachusetts.
- Landwehr, J. M., Swift, J., & Watkins, A. E. (1987). *Exploring surveys and information from samples*. Palo Alto, CA: Dale Seymour.
- Landwehr, J. M. & Watkins, A. E. (1986). *Exploring data: Quantitative literacy series*. Palo Alto, CA: Dale Seymour.
- Leavy, A. & Hourigan, M. (2016). Crime scenes and mystery players! Using driving questions to support the development of statistical literacy. *Teaching Statistics*, 38, 29–35.
- Linacre, J. M. (2016). Winsteps® (Version 3.92.0) [Computer Software]. Beaverton, Oregon: Winsteps.com. [Online: www.winsteps.com/]
- Marshall, L. & Swan, P. (2006). Using M&Ms to develop statistical literacy. *Australian Primary Mathematics Classroom*, 11(1), 15–24.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149–174.
- Ministry of Education. (2009). *The New Zealand Curriculum*. Wellington, NZ: Author.
[Online: nzcurriculum.tki.org.nz/The-New-Zealand-Curriculum]
- Moreno, J. L. (2002). Toward a statistically literacy citizenry: What statistics everyone should know. In B. Phillips (Ed.), *Proceedings of the 6th International Conference on Teaching of Statistics: Developing a Statistically Literate Society*. Voorburg, The Netherlands: International Statistical Institute.
[Online: iase-web.org/documents/papers/icots6/1b6_more.pdf]
- National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.
- Newman, C. M., Obremski, T. E., & Scheaffer, R. L. (1987). *Exploring probability*. [Quantitative Literacy Series] Palo Alto, CA: Dale Seymour.
- Petty, N. W. (2010). Creating youtube videos that engage students and enhance learning in statistics and excel. In C. Reading (Ed.), *Proceedings of the 8th International Conference on*

- Teaching Statistics: Data and Context in Statistics Education: Towards an Evidence-Based Society*. Voorburg, The Netherlands: International Association for Statistical Education. [Online: iase-web.org/documents/papers/icots8/ICOTS8_C165_PETTY.pdf]
- Petty, N. W. (2014). Taking statistical literacy to the masses with youtube, blogging, facebook and twitter. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Proceedings of the Ninth International Conference on Teaching Statistics: Sustainability in Statistics Education*. Voorburg, The Netherlands: International Association for Statistical Education. [Online: iase-web.org/icots/9/proceedings/pdfs/ICOTS9_7A3_PETTY.pdf]
- Pierce, R., Chick, H., Watson, J., Les, M., & Dalton, M. (2014). A statistical literacy hierarchy for interpreting educational system data. *Australian Journal of Education*, 58(2), 195–217. [Online: doi.org/10.1177/0004944114530067]
- Scheaffer, R. L. (2003). Statistics and quantitative literacy. In B. L. Madison & L. A. Steen (Eds.), *Quantitative literacy: Why numeracy matters for schools and colleges* (pp. 145–152). Princeton, NJ: The National Council on Education and the Disciplines.
- Scheaffer, R. L., Watkins, A. E., & Landwehr, J. M. (1998). What every high-school graduate should know about statistics. In S. P. Lajoie (Ed.), *Reflections on statistics: Learning, teaching and assessment in grades K-12* (pp. 3–31). Mahwah, NJ: Lawrence Erlbaum.
- Schild, M. (2002). Three kinds of statistical literacy: What should we teach? In B. Phillips (Ed.), *Proceedings of the 6th International Conference on Teaching of Statistics: Developing a Statistically Literate Society*. Voorburg, The Netherlands: International Statistical Institute. [Online: iase-web.org/documents/papers/icots6/1a2_schi.pdf]
- Schild, M. (2010). Assessing statistical literacy: Take CARE. In P. Bidgood, N. Hunt, & F. Jolliffe (Eds.), *Assessment methods in statistical education: An international perspective* (pp. 133–153). Chichester, UK: Wiley.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57, 1–22.
- Sproesser, U., Kuntze, S., & Engel, J. (2014). A multilevel perspective on factors influencing students' statistical literacy. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Proceedings of the Ninth International Conference on Teaching Statistics: Sustainability in Statistics Education*. Voorburg, The Netherlands: International Association for Statistical Education. [Online: iase-web.org/icots/9/proceedings/pdfs/ICOTS9_7E2_SPROESSER.pdf]
- St. Clair, K. & Chihara, L. (2012). Team-based learning in a statistical literacy class. *Journal of Statistics Education* 20(1). [Online: ww2.amstat.org/publications/jse/v20n1/chihara.pdf]
- Steen, L. A. (Ed.). (2001). *Mathematics and democracy: The case for quantitative literacy*. Washington, DC: Woodrow Wilson National Fellowship Foundation.
- Tishkovskaya, S. & Lancaster, G. A. (2012). Statistical education in the 21st century: A review of challenges, teaching innovations and strategies for reform. *Journal of Statistics Education* 20(2). [Online: ww2.amstat.org/publications/jse/v20n2/tishkovskaya.pdf]
- Tversky, A. & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgement. *Psychological Review*, 90, 293–315.
- United Nations Economic Commission for Europe (2012). *Making data meaningful. Part 4: A guide to improving statistical literacy*. Geneva: United Nations. [Online: www.unece.org/fileadmin/DAM/stats/documents/writing/Making_Data_Meaningful_Part_4_for_Web.pdf]
- Utts, J. (2003). What educated citizens should know about statistics and probability. *The American Statistician*, 57(2), 74–79. [Online: doi.org/10.1198/0003130031630]
- Vita, A. C., Santana, E., Cazorla, I., Oliveira, S. (2010). Analysis of activity at statistical literacy: Contributions of the activity theory. In C. Reading (Ed.), *Proceedings of the 8th*

- International Conference on Teaching Statistics: Data and Context in Statistics Education: Towards an Evidence-Based Society*. Voorburg, The Netherlands: International Association for Statistical Education.
[Online: iase-web.org/documents/papers/icots8/ICOTS8_C141_VITA.pdf]
- Wallman, K. K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association*, 88, No. 421, 1–8.
- Watson, J. M. (1997). Assessing statistical literacy using the media. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 107–121). Amsterdam: IOS Press & The International Statistical Institute.
- Watson, J. M. (1998). Assessment of statistical understanding in a media context. In L. Pereira-Mendoza, L. S. Kea, T. W. Kee, & W. Wong (Eds.), *Proceedings of the 5th International Conference on the Teaching of Statistics: Statistical Education – Expanding the Network*. Voorburg, The Netherlands: International Statistical Institute.
[Online: iase-web.org/documents/papers/icots5/Topic6w.pdf]
- Watson, J. M. (2006). *Statistical literacy at school: Growth and goals*. Mahwah, NJ: Lawrence Erlbaum.
- Watson, J. M. (2011). Personal and professional numeracy: A unit for pre-service teachers at the University of Tasmania. *Numeracy*, 4(1), Article 2.
[Online: doi.org/10.5038/1936-4660.4.1.2]
- Watson, J. M. & Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal*, 2(2), 3–46.
[Online: [iase-web.org/documents/SERJ/SERJ2\(2\)_Watson_Callingham.pdf](http://iase-web.org/documents/SERJ/SERJ2(2)_Watson_Callingham.pdf)]
- Watson, J. M. & Callingham, R. (2005). Statistical literacy: From idiosyncratic to critical thinking. In G. Burrill & M. Camden (Eds.), *Curricular Development in Statistics Education. International Association for Statistical Education (IASE) Roundtable, Lund, Sweden, 2004* (pp. 116–162). Voorburg, The Netherlands: International Statistical Institute.
[Online: iase-web.org/documents/papers/rt2004/4.1_Watson&Callingham.pdf]
- Watson, J. & Callingham, R. (2013). Likelihood and sample size: The understandings of students and their teachers. *Journal of Mathematical Behavior*, 32, 660–672.
- Watson, J. & Callingham, R. (2014). Two-way tables: Issues at the heart of statistics and probability for students and teachers. *Mathematical Thinking and Learning*, 16, 254–284.
- Watson, J. & Callingham, R. (2015a). Getting out of bed: Students’ beliefs. In M. Marshman, V. Geiger, & A. Bennison (Eds.), *Mathematics education in the margins* (pp. 619–626). Adelaide, SA: MERGA.
- Watson, J. & Callingham, R. (2015b). Lung disease, indigestion, and two-way tables. *Investigations in Mathematics Learning*, 8(2), 1–16.
- Watson, J. M. & Caney, A. (2005). Development of reasoning about random events. *Focus on Learning Problems in Mathematics*, 27(4), 1–42.
- Watson, J. M., Caney, A., & Kelly, B. A. (2004). Beliefs about chance in the middle years: Longitudinal change. In I. Putt, R. Faragher, & M. McLean (Eds.), *Mathematics education for the third millennium: Towards 2010* (Proceedings of the 27th Annual Conference of the Mathematics Education Research Group of Australasia, Townsville, Vol. 2, pp. 581–588). Sydney, NSW: MERGA.
- Watson, J. M. & Moritz, J. B. (2000). The longitudinal development of understanding of average. *Mathematical Thinking and Learning*, 2(1&2), 11–50.
- Watson, J. & Nathan, E. (2010a). Assessing the interpretation of two-way tables as part of statistical literacy. In C. Reading (Ed.), *Proceedings of the 8th International Conference on Teaching Statistics: Data and Context in Statistics Education: Towards an Evidence-Based Society*. Voorburg, The Netherlands: International Association for Statistical

- Education.
 [Online: iase-web.org/documents/papers/icots8/ICOTS8_5E1_WATSON.pdf]
- Watson, J. M. & Nathan, E. L. (2010b). Biased sampling and PCK: The case of the marijuana problem. In L. Sparrow, B. Kissane, & C. Hurst (Eds.), *Shaping the future of mathematics education* (Vol. 2, pp. 610–617). Fremantle, WA: MERGA.
 [Online: www.merga.net.au/documents/MERGA33_Watson&Nathan.pdf]
- Wilson, M. & Draney, K. (2000, May). Standard Mapping: A technique for setting standards and maintaining them over time. In *Models and analyses for combining and calibrating items of different types over time*. Symposium conducted at the International Conference on Measurement and Multivariate Analysis, Banff, Canada.
- Zieffler, A., Garfield, J., Alt, S., Dupuis, D., Holleque K., & Chang, B. (2008). What does research suggest about the teaching and learning of introductory statistics at the college level? A review of the literature. *Journal of Statistics Education*, 16(2).
 [Online: www.amstat.org/publications/jse/v16n2/zieffler.html]
- Ziegler, L. (2014). Reconceptualizing statistical literacy: Developing an assessment for the modern introductory statistics course. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Proceedings of the Ninth International Conference on Teaching Statistics: Sustainability in Statistics Education*. Voorburg, The Netherlands: International Association for Statistical Education.
 [Online: iase-web.org/icots/9/proceedings/pdfs/ICOTS9_7F1_ZIEGLER.pdf]

ROSEMARY CALLINGHAM
 University of Tasmania
 Locked Bag 1307, Launceston, Tasmania 7250
 Australia

APPENDIX: SURVEY ITEMS WITH RUBRICS FOR CODING

This appendix contains items used in the StatSmart project that were not used in Callingham and Watson (2005) or Watson and Callingham (2003). Rubrics were modified for some items, Ted was a reworded version of the original, and the final four items were reworded from the item reported in JMES based on a new media report. Hence for some items where appropriate two sources are acknowledged, the original source and the modified item.

RUTH (Watson & Caney, 2005). Ruth says that choosing consecutive numbers like 1, 2, 3, 4, 5, 6 increases your chances of winning Tattslotto. But Jenny says that there is a larger chance of getting a random set of numbers. She used a random number wheel to choose the following six numbers: 7, 35, 16, 26, 41, 11. Which set of numbers would you choose and why?

<i>Code</i>	<i>Description of Rubric</i>
4	Not influenced by consecutive order of numbers and ideas of variation/spread with explicit mention of all numbers having an equal chance. May include ideas on sets of numbers and chance.
3	Express a generalised idea about chance WITH qualitative chance statements.
2	Express contradictory ideas and show no recognition of contradiction OR broad chance reasoning with no qualified chance statement.
1	Influenced by consecutive order of numbers and expected variation/spread of numbers. Choose Jenny's numbers.
0	NR; Idiosyncratic, no reasoning.

TATS (Watson & Caney, 2005). One day Claire won Tattsлото 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?

<i>Code</i>	<i>Description of Rubric</i>
5	Higher reasoning about physiological states as a reason for choosing certain numbers.
4	Two or more relevant propositions made, one of which is explicit in expressing the equality of chance for all numbers or combinations of numbers.
3	All combinations of numbers have the same chance of occurring on any draw, and yet do not do so explicitly.
2	Reject 'luck' belief or state that numbers were unlikely to occur again, or less likely to occur than other numbers.
1	Affirm a belief in being lucky.
0	NR; irrelevant response; Yes or No.

T2X2 (Batanero, Estepa, Godino, & Green 1996; Watson & Callingham, 2014). The following information is from a survey about smoking and lung disease among 250 people. Using this information, do you think that for this sample of people lung disease depended on smoking? Explain your answer.

	Lung disease	No lung disease	Total
Smoking	90	60	150
No smoking	60	40	100
Total	150	100	250

<i>Code</i>	<i>Description of Rubric</i>
5	Critically examine all information (implies look at all cells) and/or correctly state proportions and percentages.
4	Agree or disagree with statement, citing evidence from the data focusing within, between or across columns (2 or 3 cells).
3	Agree or disagree with statement citing evidence from a single cell in the data or a single variable in 2 cells.
2	Critical analysis of potential survey methods and the limitations of the methods used to collect the data.
1	Agree or disagree with statement, stating more/less chance, but with no explicit reference to the data. Positive or negative statements justified with knowledge of the content area, but not based on data.
0	Positive or negative statement with no justification. No reason. No response.

SKINR (Batanero, Estepa, Godino, & Green 1996; Watson & Callingham, 2014). The following information is from a study to see if a non-active life style has some relationship with skin allergies. Using this information, do you think that people's skin allergies depend on a non-active life style? Explain your answer.

	Skin allergy	No skin allergy	Total
Non-active life	13	3	16
Active lifestyle	2	12	14
Total	15	15	30

<i>Code</i>	<i>Description of Rubric</i>
4	Suggesting an association using information from all four central cells (maybe totals as well); for example, comparison of diagonal cells (13,12) with opposite diagonal (2,3), use of ratio, percent or proportion.
3	Reference to data/information in the table using 2 cells only in suggesting there is an association.
2	Reference to data/ information in the table but unquantified in suggesting there is an association.
1	Comment based on opinion or context, no numbers or general reference to data/information in the table; or use of inappropriate numbers.
0	NR

STOMR (Batanero, Estepa, Godino, & Green 1996; Watson & Callingham, 2015b). The following information is from a study to assess if a certain drug produces indigestion (stomach trouble) in elderly people. Using this information, do you think that the elderly people's indigestion depends on taking the drug? Explain your answer.

	Indigestion	No Indigestion	Total
Drug taken	8	8	16
No drug taken	7	1	8
Total	15	9	24

<i>Code</i>	<i>Description of Rubric</i>
4	Use of ratio or percentage appropriately in relation to totals with appropriately worded conclusion.
3	Use of 2 cells for evidence to answer.
2	Use of 1 cell for evidence to answer.
1	Comment on context of question; no numbers.
0	NR, Idiosyncratic, "Yes", "No".

HOSP (Kahneman & Tversky, 1972; Watson & Callingham, 2013). A town has two hospitals. In the West hospital on one day 45 babies are born. In the East hospital 15 babies are born. In which hospital are there likely to be more than 60% boys born or is the chance the same? Explain why.

<i>Code</i>	<i>Description of Rubric</i>
3	60% of 15 is 9; 60% of 45 is 27 – only requires 3 more boys to be > 9 but 6 more boys to be > 27 [descriptive with numbers].
2	Small (East) because: Smaller samples have more variation; Larger samples are closer to expected 50% [descriptive response, no numbers].
1	Conclusion based on size: Larger (West) because more babies – more chance of boys. Same chance because 50% no matter how many born.
0	NR

TED (Kahneman & Tversky, 1972; Watson & Callingham, 2013). Ted and Jed are each tossing a fair coin. Ted tosses his 10 times and Jed tosses his 30 times. Which one of them is more likely to get more than 60% heads or do they have the same chance? Explain why.

<i>Code</i>	<i>Description of Rubric</i>
3	Ted (10) because 60% of 10 is 6 and 60% of 30 is 18 so it only takes 2 more heads than tails for Ted but 4 more for Jed [describing with numbers].
2	Ted (10) because a smaller number of tails will have (should have) more variation, a larger number of tails will be (should be) closer to 50% heads [describing without numbers].
1	Jed (30) because he has more tosses to get heads; same chance because heads are 50-50.
0	Ted / Jed because he is lucky.

F2.3 (Watson & Moritz, 2000). The average number of children in 10 families in the neighbourhood is 2.3. One family with 5 children leaves the neighbourhood. What is the average number of children per family now? Show your work here.

<i>Code</i>	<i>Description of Rubric</i>
2	Correct answer with working out shown and/or clear explanation.
1	Partial attempt, incomplete – can recognise some aspect of problem.
0	NR; incoherent work with numbers, but no sense. Correct answer – no explanation.

The following tasks (Watson & Callingham, 2015a; Watson, Caney, & Kelly, 2004) are placed in the sequence BED (alone), JMES, FENG, and PSYC (together). They share the same rubric, which is shown after PSYC.

BED. The following extract is from a newspaper. Explain what you think of these claims?

Left is the right way to exit bed

Feng shui expert Jan Cisek said getting out of the bed on the left was associated with all that people held dear – family and health, money and power. Psychologist Pete Cohen said the left side helped us to think rationally about the day ahead.

JMES. Every morning James gets out of the left side of the bed. He says that this increases his chance of getting good marks. Explain what you think of this claim.

FENG. Now consider the following newspaper headline. Explain what you think of this claim.

Left is the right way to exit bed

Feng shui expert Jan Cisek said getting out of the bed on the left was associated with all that people held dear – family and health, money and power ...

PSYC. The following headline is also in the same article. Explain what you think of this claim.

Left is the right way to exit bed

... Psychologist Pete Cohen said that getting out of bed on the left side helped us to think rationally about the day ahead.

<i>Code</i>	<i>Description of Rubric</i>
4	Combination of more than one argument, based on lack of evidence, physical conditions, and/or based on a psychological belief that may assist performance.
3	Presentation of one argument, either based on a lack of evidence, physical conditions, or based on a psychological belief that may assist performance.
2	Reject claim; simple disagreement with no justification.
1	Agree with James or Feng shui expert or Psychologist.
0	NR, Yes or No only.