INFLUENCING STATISTICAL LITERACY IN THE MIDDLE YEARS OF SCHOOLING: THE FIRST YEAR OF THE *STATSMART* PROJECT

Rosemary CALLINGHAM
University of New England, Australia

Jane M. WATSON
University of Tasmania, Australia

Julie DONNE
University of Tasmania, Australia

StatSmart is a three-year longitudinal study that started in 2007. The aim is to improve students levels of statistical literacy through changing teachers' approaches to teaching statistics in the classroom. Teacher and student change will be tracked over three years to determine factors of effective professional learning programs that impact on students' outcome. Baseline findings indicate considerable differences among groups of students, but surprising similarities among teachers.

When chance and data was introduced as a component of the school mathematics curriculum (Australian Education Council, 1991; National Council of Teachers of Mathematics, 1989), there was little research on which to base curriculum development. Over the following years a considerable body of evidence emerged about school students' development of understanding of the various concepts of chance and data, including sampling, data representation, average, chance and probability, inference, and variation (see Watson, 2006 for a summary). In 2003, the findings from several studies over ten years were consolidated into a hierarchy of statistical literacy (Watson & Callingham, 2003). This hierarchy provided a basis on which to consider students' growth in understanding over time. Despite the extensive research, however, evidence suggested that there had been little growth in students' statistical literacy over a ten-year period (Watson, Kelly, & Izard, 2006).

One reason advanced for this disappointing outcome was the relative lack of confidence of teachers in teaching chance and data (Callingham, Watson, Collis, & Moritz, 1995), with the subsequent development and refinement of a profiling instrument for teachers (Watson, 1998, 2001) that addressed Shulman's (1987) categories of teacher knowledge. Evidence suggested that positive outcomes for students could be achieved with appropriate teaching approaches (e.g., Watson & Kelly, 2004).

Against this background the *StatSmart* project was conceived. *StatSmart* is a three-year (2007–2009), Australian Research Council Linkage funded project involving primary and high schools in three states of Australia. The major educational aim of the study is to improve the statistical literacy of school students in the middle years of schooling through the improvement of the teaching of statistics at the middle and high school levels. The project will document and evaluate teachers' professional learning in terms of the connections required among the seven forms of knowledge suggested by Shulman (1987): content knowledge, general pedagogical knowledge, curriculum knowledge, pedagogical content knowledge, knowledge of learners and their characteristics, knowledge of education contexts, and knowledge of education ends, purposes, and values using an adaptation of Watson's (2001) profiling instrument. In addition it will track the growth of students' understanding, using surveys comprising items from statistical literacy instruments used in earlier studies (Callingham & Watson, 2005; Watson & Callingham, 2003). The research design is complex, using multiple, linked survey forms for students with three data points, and a single teacher profiling instrument, also with three data points (Callingham & Watson, 2007). In addition, interviews with students and teachers will provide additional details not available from surveys.

During the first year of the project teachers received initial professional development at a conference funded by the Australian Bureau of Statistics, one of the project partners. Another partner, Key Curriculum Press, supplied *TinkerPlots* software (Konold & Miller, 2005) and associated training. Some additional professional development was also undertaken with most teachers.

METHOD

Teachers and students across three states and 19 schools completed initial survey instruments to provide baseline data for the project. The students were in grades 5 to 10, although not every state provided all grades. The breakdown of the student sample is shown in Table 1. Of the 1205 students, 52.9 percent (*n*=638) were male and 47.1 percent (*n*=567) were female. Of the 19 schools involved, five were single-sex (2 all male; 3 all female); three were primary only, seven were high schools and nine included both primary and high school grades. In every state, schools were from government, Catholic and independent systems.

Table 1
Student sample

	State A	State B	State C	Total
Gr5	0	29	12	41
Gr6	35	125	15	175
Gr7	90	155	187	432
Gr8	95	177	52	324
Gr9	64	0	73	137
Gr10	0	0	96	96
Total	284	486	435	1205

Altogether 42 teachers completed the profiling instrument, evenly distributed across the three states. Of these, 20 were female and 22 male. These teachers were very experienced with over 50 percent having taught for more than 15 years, and only seven having less than 5 years experience. Table 2 shows the distribution of experience by state

Table 2
Teacher experience by state

	State A	State B	State C	Total
First year	0	0	1	1
2-5yrs	4	1	1	6
6-15yrs	5	5	2	12
16-25yrs	1	4	6	11
>25yrs	4	4	4	12
Total	14	14	14	42

Teachers were also asked to indicate what was the highest level of mathematics that they taught, and their mathematics background. Table 3 shows the highest level of mathematics taught against the reported mathematics background for the 39 teachers who provided both bits of information. Of interest is the fact that only about one-quarter of the teachers had a mathematics major, even when they taught in the higher grades. Teachers were not specifically asked about any statistics training that they had received.

Table 3
Teacher mathematics background and highest level of mathematics taught

	No maths	1 semester	1 year	Maths major	Total
Primary	3	2	2	0	7
Middle	1	0	2	1	4
Junior Secondary (Grades 7/8 – 10)	1	3	3	1	8
Senior Secondary (Grades 11/12)	1	0	10	9	20
Total	6	5	17	11	39

Early in 2007 and prior to any teaching of statistics, students completed one of three test forms, all linked through common items so that by using Rasch measurement approaches (Bond & Fox, 2007) direct comparisons could be made within and between groups. Each test was composed of between 24 and 29 items covering all facets of statistical literacy. Responses were coded using rubrics from previous administrations of the items so that consistency of interpretation and judgement was ensured. The scale was checked for fit to the model and student ability estimates obtained in logits (the unit of Rasch measurement). These estimates were used to make comparisons between sub-groups of students by grade, including gender and state, so that curriculum effects could be determined.

The process of determining levels was carried out in two stages. The items were calibrated using the Quest computer program (Adams & Khoo, 1996), anchored to a previous administration of tests of statistical literacy through a set of common items to establish the scale (see Griffin & Callingham, 2006 for details about this process). By anchoring to a previous administration of a set of items, the levels of statistical literacy on the Watson and Callingham (2003) hierarchy were identified using the logit values previously determined. Logits are the units of Rasch measurement. Second the students' ability measures, also in logits, were estimated against the same anchored scale. Students were then grouped into the levels of the hierarchy based on these estimated values. The proportion of students by grade in each level provided baseline data for the study.

Teachers' responses to the profiling instrument were coded using previously used rubrics. The items were calibrated to identify a single scale of teacher knowledge for teaching statistics (Watson, Callingham, & Donne, in press). In addition several subscales were identified based on components of the profile: confidence in teaching statistics topics; beliefs about statistics; teaching strategies; assessment strategies; and pedagogical content knowledge. Rasch estimates of teacher ability were obtained for the overall scale and for each subscale, in order to consider factors that might impact on teachers' capacity to teach statistics. These ability estimates were used to consider within and between group differences.

INITIAL FINDINGS

Student findings

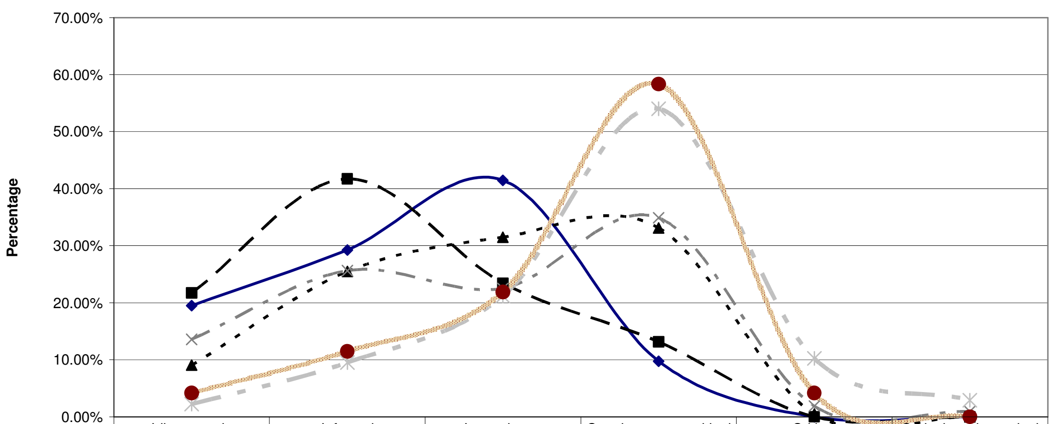
Student ability levels from the Rasch measurement estimations were grouped into levels of the statistical literacy hierarchy to provide baseline information for the study. The levels are described in Table 4.

Table 4

Description of Levels of Statistical Literacy Hierarchy (Watson & Callingham, 2003)

Level	Brief characterization of step levels of tasks
6. Critical Mathematical	Task-steps at this level demand 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	Task-steps require 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	Task-steps require 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. Inconsistent	Task-steps at this level, often in supportive formats, expect selective engagement with context, appropriate recognition of conclusions but without justification, and qualitative rather than quantitative use of statistical ideas.
2. Informal	Task-steps require 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	Task-steps at this level suggest idiosyncratic engagement with context, tautological use of terminology, and basic mathematical skills associated with one-to-one counting and reading cell values in tables.

Figure 2 shows the percentage of students in the identified levels of the statistical literacy hierarchy. It is noticeable that after grade 6, a majority of students is in the consistent non-critical level. At this level, students display appropriate but non-critical engagement with context, and some facility with means, graphing and simple probability. Students do not appear to progress much beyond this level. Grade 7 shows the most spread in the distribution across the levels, and this may reflect the fact that in two states grade 7 is in high school settings, whereas in the third it is in a primary setting. In this sample, grade 5 students appear somewhat more competent than grade 6, but this may be an artefact of the sample because there are relatively few grade 5 students involved.



	Idiosyncratic	Informal	Inconsistent	Consistent non-critical	Critical	Critical-mathematical
G5	19.51%	29.27%	41.46%	9.76%	0.00%	0.00%
G6	21.71%	41.71%	23.43%	13.14%	0.00%	0.00%
G7	9.03%	25.46%	31.48%	33.10%	0.46%	0.00%
G8	13.58%	25.62%	22.53%	34.88%	1.85%	0.93%
G9	2.19%	9.49%	21.17%	54.01%	10.22%	2.92%
G10	4.17%	11.46%	21.88%	58.33%	4.17%	0.00%

Statistical literacy levels

G5 G6 G7 G8 G9 G10

Figure 1. Students' levels of statistical literacy by grade.

Students in different grades were also compared across states. Figure 2 shows mean ability measures by grade for all three states. Error bars are left off for clarity. These findings suggest that there are considerable differences across states, possibly reflecting varying curriculum emphases. These differences were all statistically significant using *t*-tests (Grade 5, df = 39, t = 4.167; Grade 9, df = 135, t = 2.934) or one-way ANOVA (Grade 6, df = 2, F = 13.766; Grade 7, df = 2, F = 12.692; Grade 8, df = 2, F = 11.029). Clearly at the start of the *StatSmart* project there is considerable disparity across the three states involved.

It is also noticeable that there does not appear to be any fall off in performance between the primary and high school settings. A plateau or drop in performance has been reported in the middle years of schooling in a number of studies (e.g., Hill, Rowe, Holmes-Smith, & Russell, 1996) and this is usually attributed to the transition from the informal primary setting to the greater formality in high schools. Among the *StatSmart* students, however, there is a plateau or drop in performance between the first and second year of high school, that is from grade 7 to grade 8 in states B and C, and from grade 8 to grade 9 in state A. There is a drop in performance between grade 5 and grade 6 in state B; however this may be due to the small number of grade 5 students (n = 29) who mainly came from one highly regarded school, whereas nearly half of the grade 6 students came from a school in a very low socio-economic area that was generally regarded as an academically weaker school.

Gender groupings within grade were examined for differences. Significant differences were found in grade 7 in favour of boys (df = 428, t = 2.610) and grade 8 in favour of girls (df = 320, t = -4.312). There were no other significant differences along gender lines. Because only 119 students spoke a language other than English, and these students were spread among the grades, there was no attempt to consider language background as a variable.

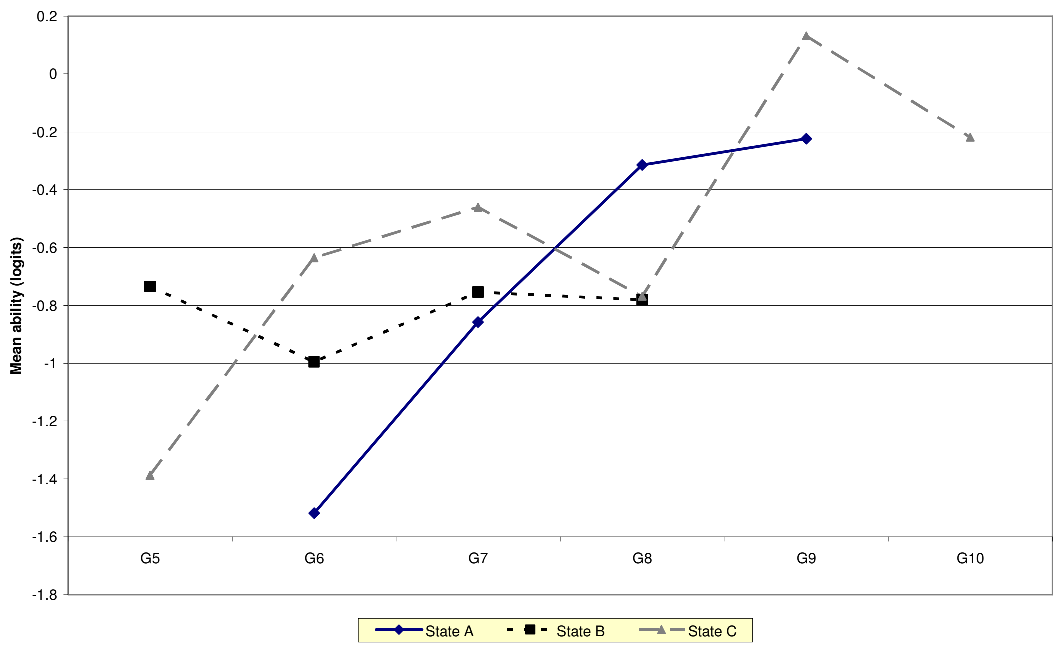


Figure 2. Mean ability estimates by grade for three states.

Teacher findings

One-way ANOVA tests were undertaken using state, teaching experience, mathematics background and highest level of mathematics taught with each of the scales of teacher knowledge: confidence, beliefs, pedagogical content knowledge, teaching approaches, assessment approaches and the overall scale. Of these tests only the confidence scale by highest level of mathematics taught (df = 3; F = 5.1) and by state (df = 3; F = 4.356), and the overall scale by highest level of mathematics taught (df = 3; F = 3.202) showed statistical significance. The teachers in State C mostly taught students in Years 11 and 12, in addition to those in the lower grades of secondary school, and there were no primary schools represented in this state. These factors may have contributed to the findings. In general, however, the teachers were a very homogeneous group.

With respect to the scales, teachers were least confident about teaching ideas about variation, sampling, inference, prediction, connecting statistics to other learning areas and statistical literacy in the media. In contrast they were more likely to express high confidence in teaching familiar curriculum topics such as mean, median, mode and histograms. This finding was similar in the beliefs scale, where teachers found it easier to agree that they could explain average to someone else, but harder to agree that they used statistics in decision making, that it was possible to use the same statistics to argue both sides of a case, and that they could explain how an opinion poll worked. On the attitudinal scales, it seemed that the more familiar the idea targeted by the item, the easier it was for teachers to endorse.

On the scale addressing teachers' pedagogical content knowledge (PCK), in general items that asked teachers to suggest likely responses that their student would make were easier than items where teachers were asked how they would intervene to develop their students' understanding. The most difficult PCK items were the two asking teachers how they would address students' inappropriate responses to drawing a sketch showing association between two variables, and the inappropriate questions students might ask. It seems that teachers are more able to imagine what their own students are likely to do than to create suitable responses to address inappropriate student answers.

In terms of teaching strategies, these teachers appeared to be relatively conventional, with 90.2 percent (37 of 41 respondents) reporting use of drill-and-practice at least once a week. At least weekly use of exposition methods ("chalk-and-talk") was reported by half of the respondents, whereas all teachers who responded to the question reported using concrete materials only occasionally, with 75 percent using them less than once a year. Given that a number of teachers were teaching primary grades this low level of use of concrete materials was surprising. Teachers' reported assessment processes were less conventional, however. The most frequently used assessment process was a short investigation with over half of all teachers (23 of 42 respondents) using these once a week, and two additional teachers reporting that they used them every lesson. Other common assessment processes were one-to-one discussion and individual questioning of students, both used at least once a week by a majority of teachers. All teachers used teacher designed tests once a term or once a year, as might be expected.

DISCUSSION

The results presented are the first from a longitudinal three-year study in which students and their teachers will be tracked over time. The aim is to influence students' outcomes by identifying aspects of teachers' knowledge that contribute to these. Initial findings from the *StatSmart* project indicate diversity among students but surprising similarity among teachers. Given the variety of teachers' experience and mathematical background, as well the range of contexts in which they were teaching, more differences might have been anticipated. It is possible that interview data to be collected this year could provide finer detail about the manner in which teachers work, and some explanation for the unanticipated similarities.

The variation in students across the grades of schooling was anticipated. In some grades, however, the numbers of students are too small to claim with confidence that the proportions of students in the levels of the statistical literacy hierarchy provide definitive benchmarks for the general population. Nevertheless, these early findings do suggest that students are not achieving the highest levels of statistical understanding. It is one of the aims of the *StatSmart* project to increase the proportion of students in the Critical and Critical-Mathematical levels. No analysis by type of school has yet been possible, because the number of students in any one grade in a particular type of school has been too small. The variety of school types, from a range of socio-economic backgrounds would lead to a prediction of differential student achievement.

The state differences observed are likely to be due to curriculum effects. The mathematics curriculum operates within a different framework in each state. Although the content is similar, the context in which it is delivered, such as in which grade a particular topic is introduced, the approach to teaching and the nature of the expected outcomes against which teachers must report, is varied from state to state. A curriculum mapping exercise to analyse the state documents could reveal some explanations for the observed student differences. A curriculum-based explanation might also account for the pattern of achievement across the grades, with a plateau effect occurring between the first and second years of high school. Further analysis, especially growth over time as individual students are tracked across the years of the study, should provide further insights.

These initial results indicate that, at least as far as statistical literacy is concerned, the teacher group is starting from a comparable base regardless of the background variables considered. Students, however, are entering the study with obvious differences, not only age-related. These are likely to be due to prior experiences, and interviews with students during the second year should reveal more information. At this beginning stage of *StatSmart*, it is not possible to establish a relationship between teacher variables and students' outcomes because the teachers had not taught their students previously. The second wave of data collection, an end-of-year survey of students linked through common items to the original baseline survey, will provide information about students' growth over time. This change will be linked to teacher and school variables so that explanatory models can be built to consider the impact on students' outcomes. The first end-of-year survey data are being coded and entered so that initial results can be presented at the *StatSmart* teachers' conference in May 2008.

The scale of the project, across three states, 19 schools and over 40 teachers, provides considerable challenges. Data are being managed using a relational database so that teachers can be reliably linked to their students, and these connections can be tracked over time as students move in and out of *StatSmart* classes. Rather than attempt to set up a controlled study, the project aims to collect data under "real world" educational conditions. Already there are issues about maintaining impetus and interest among teachers. Although teachers are eager to be involved, the reality of the situation is that some teachers have already left the project to be replaced by new teachers from their schools, some are finding it difficult to teach the required statistics unit and some have been unable to participate in additional professional learning. In some schools, participation in CensusAtSchool is problematic. This activity is run by the Australian Bureau of Statistics and involves collecting data from individual students, which they then use in a variety of ways to develop statistical understanding, so that CensusAtSchool is an essential component of the project. All schools have been given *TinkerPlots* software (Konold & Miller, 2005) and different approaches to integrating its use in schools are already being identified. Some schools are planning cross-curriculum integrated units, predicated on using the software to explore real data sets in a variety of settings. Other schools are using *TinkerPlots* only in mathematics classes. These diverse considerations are likely to have an impact on both teachers' professional learning and students' outcomes. Although some situations were anticipated, others have emerged, and in the second year of the study a systematic documentation of such happenings will become important.

The baseline data from *StatSmart* hint at potentially challenging and useful findings yet to come. Over the next two years, additional data will be collected from two new cohorts of students, and the original students will be tracked across time. Teachers will also be monitored to identify changes. These data should provide information to policy makers, systems and teachers about effective approaches to developing statistical literacy.

ACKNOWLEDGEMENT

The *StatSmart* study research was funded by Australian Research Council grant number LP0669106, with support from the Australian Bureau of Statistics, Key Curriculum Press, and the Baker Centre for School Mathematics.

REFERENCES

Adams, R. J., & Khoo, S. T. (1996). *Quest: Interactive item analysis system. Version 2.1* [Computer software]. Melbourne: Australian Council for Educational Research.

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

Bond, T.G. & Fox, C.M. (2007). *Applying the Rasch model. Fundamental measurement in the human sciences* (2nd. Ed.) Mahwah, NJ: Lawrence Erlbaum.

Callingham, R. & Watson, J. M. (2005). Measuring statistical literacy. *Journal of Applied Measurement, 6* (1), 29, 19-47.

Callingham, R., & Watson, J. (2008). Overcoming research design issues using Rasch measurement: The *StatSmart* project. In P. Jeffery (Ed.) *Proceedings of the annual conference of the Australian Association for Research in Education, Perth.* Retrieved 9 March 2008 from http://www.aare.edu.au/07pap/cal07042.pdf

Callingham, R.A., Watson, J.M., Collis, K.F., & Moritz, J.B. (1995). Teacher attitudes towards chance and data. In B. Atweh & S. Flavel (Eds.), *Galtha* (Proceedings of the 18th Annual Conference of the Mathematics Education Research Group of Australasia, pp. 143-150). Darwin, NT: MERGA.

Griffin, P. & Callingham, R. (2006). A twenty-year study of mathematics achievement. *Journal for Research in Mathematics Education, 37* (3), 167-186.

Hill, P.W., Rowe, K.J., Holmes-Smith, P., & Russell, V.J. (1996). *The Victorian Quality Schools Project: A study of school and teacher effectiveness. Report (Volume 1*). Melbourne: Centre for Applied Educational Research, University of Melbourne.

Konold, C. & Miller, C.D. (2005). *Tinkerplots: Dynamic data exploration* [Computer software]. Emeryville, CA: Key Curriculum Press.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

Shulman, L.S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*, 1-22.

Watson, J.M. (1998). Professional development for teachers of probability and statistics: Into an era of technology. *International Statistical Review, 66*, 271-289.

Watson, J.M. (2001). Profiling teachers' competence and confidence to teach particular mathematics topics: The case of chance and data. *Journal of Mathematics Teacher Education, 4*, 305-337.

Watson, J.M. (2006). *Statistical literacy at school. Growth and goals*. Mahwah, NJ: Lawrence Erlbaum Associates.

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

Watson, J.M. & Kelly, B.A. (2004). Statistical variation in a chance setting: A two-year study. *Educational Studies in Mathematics, 57*, 121-144.

Watson, J.M., Kelly, B.A., & Izard, J. F. (2006). A longitudinal study of student understanding of chance and data. *Mathematics Education Research Journal*, *18*(2), 40-55.