

DEVELOPING STATISTICAL LITERACY IN THE FINAL SCHOOL YEAR

STEPHANIE BUDGETT

*University of Auckland
s.budgett@auckland.ac.nz*

DRUSILLA ROSE

*Westlake Girls' School
DRose@westlakegirls.school.nz*

ABSTRACT

Statistical information pervades everyday life in the twenty-first century. Research shows, however, that the skills needed to be able to understand and critically evaluate statistical information must be specifically taught. In 2013, an externally assessed National Certificate in Educational Achievement standard in statistical literacy was introduced for the first time in New Zealand. A small exploratory study investigated a possible teaching approach designed to enable Year-13 students (aged 17-18) to critically evaluate media reports. Findings suggest that the learning trajectory required several key components including media reports as both a motivational and conceptual development tool. In addition, computer visualizations and procedural scaffolds appeared valuable tools for facilitating conceptual understanding of the margin of error.

Keywords: *Secondary school students; Statistical literacy; Margin of error*

1. INTRODUCTION

It is more than twenty years since Wallman's (1993) presidential address to the American Statistical Association entitled "Enhancing statistical literacy: enriching our society." She made a plea for statistical literacy education for all citizens in order that they may acquire the knowledge and skills needed to understand and critically evaluate the statistical information that permeates everyday life in the modern era. However, it is only in the last three years, with the realignment of the mathematics and statistics standards for the national qualification (National Certificate in Educational Achievement, NCEA) in accordance with the revised New Zealand Curriculum (Ministry of Education, 2007), that statistical literacy has begun to be taught in New Zealand classrooms. Even though the 1992 New Zealand Curriculum stated that students should evaluate statistics in the media, this was never assessed in any external examination. As a result, statistical literacy was largely ignored by the current generation of secondary mathematics and statistics teachers, who lacked both the pedagogy and the confidence needed to implement a statistical literacy curriculum (Merriman, 2005).

Implementation of the standards realignment began in 2011 with NCEA Level-1 candidates (aged 15-16 years). In terms of statistical literacy components, the externally assessed 'Demonstrate understanding of chance and data' standard (worth four credits), which is taken by the majority of Level-1 mathematics and statistics candidates,

encompasses several achievement objectives including one on *evaluating statistical investigations of probability activities undertaken by others*. However, at Level 2, the statistical literacy standard ‘Evaluate a statistically based report’ is an internally assessed standard (worth two credits). In 2012, the first year in which this Level-2 standard was offered and the year in which this study was undertaken, the uptake was very low; only 12.8% of all mathematics and statistics candidates, generally those students embarking on a less academic pathway with more of a focus on internally assessed standards. In 2013 there will be seven new Level-3 standards in the mathematics and statistics subject area, four internally assessed standards and three externally assessed standards. Schools are expected to offer five or six of these standards to suit the needs of their students. The Level-3 statistical literacy standard ‘Evaluate statistically based reports’ (four credits) is to be externally assessed and it is hoped that this will encourage a high uptake.

The new Level-3 statistical literacy standard requires students (17-18 year olds) to evaluate statistically-based reports, including surveys, polls, experiments and observational studies. Evaluation of polls and surveys requires an understanding of both sampling and non-sampling errors. However, research on how to teach statistical literacy is currently very limited. Therefore, research into possible teaching approaches at the final school year level, Year 13, is required. The main purpose of the study that provided the motivation for this paper was to investigate a possible approach to the teaching and learning of statistical literacy in statistics students with a particular focus on developing the ability to critically evaluate media reports on polls and surveys (Rose, 2013). One outcome of the study was a specifically designed hypothetical learning trajectory (HLT) for statistical literacy. Part of the resulting HLT focused on developing the concept of margin of error and to test or make a claim regarding a single proportion. The aim of this paper is to investigate the following research question: What elements of the hypothetical learning trajectory will facilitate students’ conceptual access to ideas underpinning margin of error?

2. LITERATURE REVIEW

2.1. STATISTICAL LITERACY

“... statistical literacy is becoming fundamental for living in a full democracy” (Biggeri & Zuliani, 1999, p. 2). Evident in this statement is the fact that in order to function effectively in the information-laden twenty-first century, we need to be statistically-literate. However, statistical literacy is a difficult construct to define, resulting in a diversity of views in the literature about its meaning. Part of the problem is the audience to whom statistical literacy applies.

Gal (2002) identifies two types of people: data consumers, encompassing the vast majority of people who encounter data in their daily lives, and data producers, incorporating students and researchers who are involved in empirical investigation of data. Wallman (1993) defines statistical literacy as “the ability to understand and critically evaluate statistical results that permeate our daily lives, coupled with the ability to appreciate the contributions that statistical thinking can make in public and private, professional and personal decisions” (p. 1). Gal (2002) extends this definition by including the “ability to discuss or communicate our reactions to such statistical information” (p. 3). He views statistical literacy as embedded within a broader framework of general literacy, graphical literacy and quantitative literacy and proposes a two-component model of statistical literacy: a knowledge component comprising five elements and a dispositional component involving two elements. The knowledge

component contends that the understanding and interpreting of statistical information requires general literacy and numeracy skills and the ability to interpret graphs and tables in addition to statistical knowledge including statistical methods used by data producers (Schiold, 2002). The dispositional component contends that to be statistically literate a person needs to have developed a willingness to adopt a critical stance and a belief that it is legitimate to be critical of statistical messages.

In defining statistical literacy as a goal for students undertaking a tertiary introductory statistics course, Rumsey (2002a) defined two levels of statistical literacy: statistical competence and statistical citizenship. She views statistical citizenship as “the ultimate goal of developing the ability to function as an educated person in today’s age of information [which] may very well require high order statistical reasoning and thinking” (p. 3). A key element in Wallman’s, Gal’s and Rumsey’s definitions of statistical literacy is the ability to critically evaluate statistical information. While some people tend to believe everything that they read or hear in media reports (e.g., Blastland & Dilnot, 2010), and others will only consider information to be reliable when it is consistent with their own personal beliefs (Everitt, 1999), the ability to adopt a critical stance is important. In order to make an educated judgment there is a need to be able to assimilate statistical information in a logical manner that is free from unintentional biases. To this end, Gal (2002) has developed a list of “worry” questions which he proposes all data consumers should ask themselves whenever they encounter statistical information.

While there is a general consensus that statistical literacy involves the ability to read, understand and critically evaluate statistical information encountered during the course of everyday life, no definition that encompasses all aspects of the associated complexities has been proposed. What is certain, however, is that in order to attain statistical citizenship, a person will not only need a sound knowledge base involving statistical and general literacy and numeracy skills, but also some higher-order statistical reasoning and thinking skills.

Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz, and Woloshin (2007) define collective statistical illiteracy as “the widespread inability to understand the meaning of numbers” (p. 53). As evidenced by many (e.g., Gigerenzer et al., 2007; Goldin, 2010; Noll, 2007; Ridgway & Ridgway, 2011), it would appear that statistical illiteracy may be as widespread among today’s data communicators, including those in the professions, as it is among ordinary citizens. Statistical illiteracy is also widespread in the media (Goldin, 2010). Because the general public, as data consumers, tend to rely on the media for information, the consequence may be that of an ill-informed public on issues that have direct relevance in their lives.

It is arguable that statistical literacy is not being acquired as part of a general education through the curriculum as it is currently delivered. In fact, as demonstrated by Noll (2007), neither is it acquired by students taking specialist courses in statistics at university level. Therefore, it seems advisable that statistical literacy be given particular emphasis within existing standard statistics courses, or be the focus of a specialised course. It is a life-skill needed by everyone in order to make sense of the statistical information they come across in the course of their everyday lives. Given that this skill is not generally being acquired by students, either through secondary or tertiary curricula, statistical literacy as an additional cognitive skill requires a further course, or a substantial part of a course, centred on the critical evaluation of media reports (Schiold, 2002; Watson, 2006).

2.2. STATISTICAL LITERACY IN THE CURRICULUM

Moore (1998) and Schield (2004) argue that the core ideas of statistics, and statistical literacy in particular, are so general and fundamental and so omnipresent in modern life that they belong in an interdisciplinary curriculum in a liberal arts programme. Watson (2002) suggests that the need for units of work involving statistical literacy might act as a catalyst for building a truly integrated curriculum. She also suggests that media articles describing statistical studies have a vital role in providing an interest factor to initiate investigation and a linking factor connecting the concepts studied in the chance and data curriculum to decision-making in everyday life (Watson, 2006). In the United Kingdom, the Smith Report (2004) suggested removing much of the study of data and chance out of the post-14 secondary mathematics curriculum into other subject areas such as geography and biology.

Opposing this view, Ridgway, Nicholson, and McCusker (2007) claim that there is an urgent need to make space in the curriculum for “reasoning from evidence in realistic situations” (p. 48) and argue that teachers would need a strong background in statistics to do this. They point to “considerable logistical problems in providing a coherent set of experiences for students in statistics, if teaching takes place in a number of different curriculum areas” (p. 44). Schield (2004), also not in favour of splitting teaching across departments, cites Best’s (2004) view that the lack of a departmental owner results in critical statistical thinking becoming everyone’s responsibility and hence no-one’s.

The Guidelines for Assessment in Statistical Education (GAISE, Franklin, et al., 2007), PreK-12 Report in the USA, have a laudable goal of producing a statistically literate high-school graduate who “will know how to interpret data in the morning newspaper and will ask the right questions about statistical claims” (p. 3). However, with the report focusing on posing questions and collecting and analysing data rather than interpreting data encountered in the media or in research studies, the concepts and activities described are more suited to the development of statistical reasoning and thinking than for statistical literacy.

In contrast, the New Zealand Curriculum (Ministry of Education, 2007) separates out statistical investigations (data production) from statistical literacy (data consumption). With the emphasis on interpreting statements made by others, the progression of statistical literacy is mapped out from the beginning of formal schooling in primary school through to the end of high school. However, until formal assessment was introduced in 2012, statistical literacy was largely ignored by secondary mathematics teachers. With the advent of a formal assessment, it is now time for research into how students’ statistical literacy can be developed.

Part of statistical literacy is interpreting the margin of error, the focus of this paper. As Franklin et al. (2007, p. 1) said, “The statistically literate citizen should understand the behavior of ‘random’ samples and be able to interpret a ‘margin of sampling error’”. An awareness of the margin of error is very important for both data communicators and data consumers. The tendency of journalists to argue with point estimates ignores the fact that polls and surveys are carried out on a sample of the population of interest, and that uncertainty is a given.

The headline of the abbreviated media article in Figure 1 makes a claim that combined support for the Labour party and the Green party (45%) is higher than support for the National party (43%). However, when the margin of error (3.4%) is taken into account, this claim is not justified.

New poll has Labour-Greens ahead of National, before Chinese sales claims

Stuff.co.nz, July 17, 2015

A new opinion poll has support for the Government plunging, behind a potential Labour-Green coalition for the first time since September's election. The latest in the often volatile Roy Morgan poll showed support for National dropping 6.5 points to 43 per cent, while the combined support for Labour and the Greens rose 6 points to 45 per cent.

Figure 1. Abbreviated media article reporting on a political poll

Furthermore, although newspaper articles reporting on the results of a poll or survey will often quote the margin of error, people are not generally aware that this does not apply to subgroups, nor does it apply to percentages outside the range of 30% to 70%. There seems to be very little research on people's understanding of the margin of error, what it means, and when it does and does not apply (Budgett & Pfannkuch, 2010).

3. METHODOLOGY

The methodological approach adopted for this study was design research. Design research uses cycles of design, test and improvement. Cobb and Gravemeijer (2008) define design research as "a family of methodological approaches in which instructional design and research are interdependent" (p. 68) and which enable the simultaneous investigation of both the process of learning and the means by which it can be supported. Therefore this approach was considered particularly suitable for this study on how Year-13 students' statistical literacy might be developed.

Because statistical literacy was a topic new to both teachers and students, the design included two cycles of test and improvement. The second author conducted the two studies and collected the data. Following development of a preliminary hypothetical learning trajectory (HLT), the first cycle involved four statistics teachers which informed modifications to the HLT. The second cycle involved a small-scale teaching intervention with the second author researcher and a group of eight students.

3.1. PARTICIPANTS, DATA COLLECTION AND ANALYSIS

The four teachers involved in the first cycle were experienced and brought a diversity of background to the development of the HLT and teaching materials. They participated in two four-hour workshop sessions. The HLT, software tools, curricular materials and the pre and post-tests were critiqued and revised in accordance with feedback from the teachers (Rose, 2013). The participating school for the second cycle was a large, state, high socio-economic single sex girls' secondary school. The eight student participants were a purposive sample from Year 12 (aged 16-17 years). Three year 12 mathematics classes were randomly selected from a total of 9. From these three classes, eight students volunteered to participate. In order to minimize any disruption to the students' progress in class, it was necessary to conduct the trial in the students' own time. This resulted in some students failing to complete the trial. All eight students sat the pre-test and four (S1, S2, S3 and S4) had an in-depth interview about their responses to the pre-test. However, only seven students were present for the first teaching session and only four students attended the second teaching session. Consequently, four students sat both the pre-test and the post-test and these four (S1, S2, S3 and S5) were interviewed in-depth about their responses to the post-test. Although the focus of this study was the development of statistical reasoning in Year-13 students, Year-12 students were used since this is the cohort that was being

taught the new curriculum and would therefore have the pre-requisite knowledge base, which included completion of a Year-12 unit on informal inference.

The teaching sequence was trialed over two sessions, each lasting about three and a half hours. A 30 minute pre-test was administered prior to the first teaching session and repeated after completion of the second teaching session (see Appendix B). The interviews and teaching sessions were video-taped and transcribed. A pragmatic approach was used for the quantitative analysis of the pre- and post-tests with the assessment schedule modeled on that used by the New Zealand Qualification Authority (NZQA) which, in turn, is based on an assessment schedule adapted from Biggs and Collis' (1982) SOLO taxonomy. Qualitative analysis of the in-depth interviews was conducted with the aim of identifying critical and salient features of the students' thinking and reasoning.

3.2. HYPOTHETICAL LEARNING TRAJECTORY

Reflective analysis and synthesis of the literature in the statistical literacy arena informed the construction of a framework for the design of an HLT for developing Year-13 students' ability to critically evaluate polls and surveys in the media. The learning trajectory was devised with the aims of expanding students' existing knowledge base in terms of both general literacy and statistical concepts, providing a motivational element, and guiding students in developing a critical stance. The resulting HLT comprised twelve key components, each including several elements (Rose, 2013). These components were: (1) a motivational aspect; (2) a test of prior knowledge; (3) literacy support; (4) extension of students' current statistical and contextual knowledge base; (5) a collaborative community of enquiry; (6) image and story support; (7) development of a critical stance; (8) the use of "worry" questions; (9) development of an appreciation of non-sampling errors; (10) development of an understanding of margin of error; (11) the use of post-reading strategies; and (12) development of students' ability to communicate in writing a critical evaluation of a media report.

The motivational component was centred on hot topics in the media and linked with images and stories providing additional support to help students move from familiar to less familiar contexts. Students' existing knowledge of sampling variation needed to be explored and extended to encompass the new concepts of non-sampling error, sampling error and confidence intervals for sample proportions. Literacy support was also required in the form of pre-, while- and post-reading strategies. The use of "worry" questions, procedural scaffolds and teacher-modelling helped students to develop a critical stance and to write reports evaluating claims made in the media. Media articles and press releases, together with the claims stated in them, formed an integral component of the HLT. They were selected carefully, with consideration given to their potential for providing a motivational element; developing specific ideas and concepts; and the desired progression from familiar to less familiar contexts.

In this paper we report on the development of the components of the HLT devoted to sampling error, margin of error and testing claims for single proportions. The materials developed comprised a media article, animations, a concrete hands-on activity and computer visualization software tools. The components of the HLT were designed with regard to pedagogical theories with the aim of optimizing knowledge transfer. For example, in accordance with cognitive load theory (Van Merriënboer & Sweller, 2005), attention was given to ensuring that pre-requisite concepts were identified and revised prior to introducing new concepts, and that procedural scaffolding was provided where appropriate. In addition, hands-on activities were implemented prior to the introduction of software tools (Rossman, 2008). Furthermore, in line with dual coding theory (Clark &

Paivio, 1991), opportunities were sought to simultaneously integrate visual cues with verbal commentary.

To provide a clearer focus, the components relating specifically to margin of error are shown in Figure 2 (this framework is adapted from Rose, 2013). However, it should be noted that these components link in with other components of the HLT such as those involving non-sampling errors and those involved with development of pre-reading and while-reading strategies.

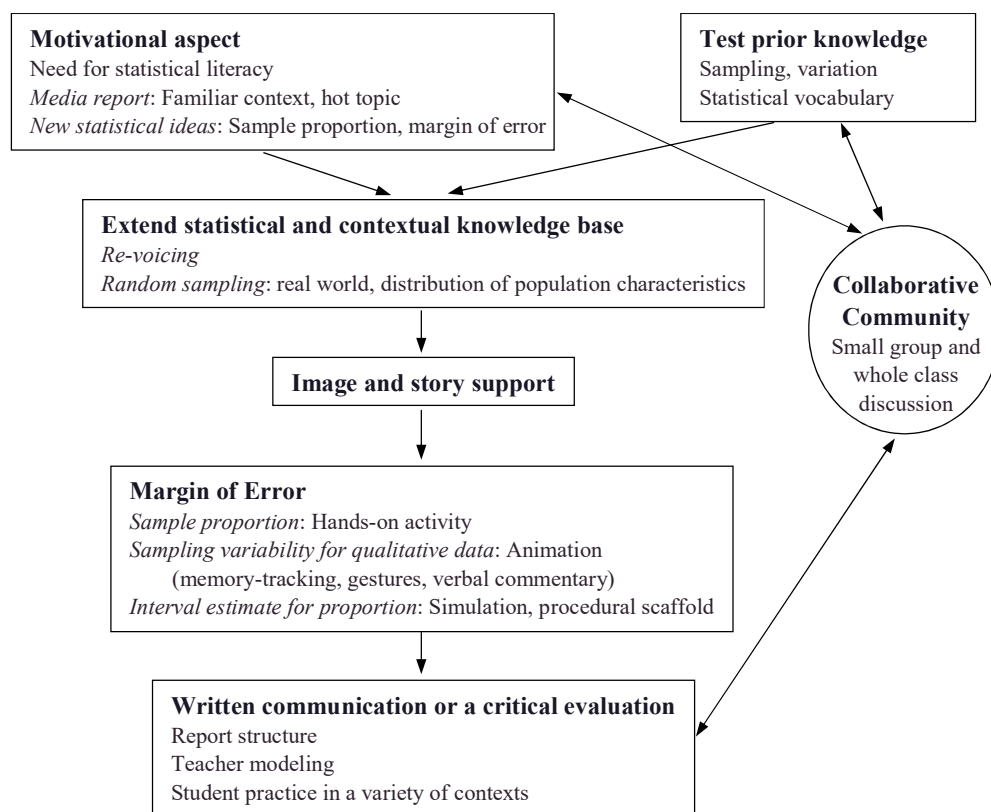


Figure 2. Margin-of-error framework within a larger hypothetical learning trajectory for developing an ability to critically evaluate polls and surveys

Establishing the statistical knowledge base Because polls and surveys were to be an integral part of the teaching intervention, the first session began with a quick class poll in order to test prior knowledge. This activity was designed to establish the notion of a poll, to introduce students to sample proportions and poll percentages and to explore, through discussion, students' current statistical knowledge of target population, sampling and inference. An invalid inferential statement was used to provoke discussion. Where appropriate, opportunities were taken to rephrase student responses with correct formal statistical terminology, thus providing vocabulary support through re-voicing.

Introducing the margin of error The selected article (see Appendix A), had a clearly stated margin of error and the headline made a claim involving a single proportion. This article therefore provided a real-life purpose for developing a conceptual understanding of margin of error, in addition to introducing students to a less familiar context.

Intuitive imagery development for margin of error While the students could be expected to be familiar with sampling variability in small samples of size 30 for measurement data, animations were selected to demonstrate the high variability in categorical data for samples of size 30, and hence the use of larger samples of sizes of 500 or 1000 in polls in the media (Wild, 2009; Wild, Pfannkuch, Regan, & Horton, 2011). The animations allow the decreasing variability of sample proportions as sample size increases ($n = 30, 100, 300,$ and 1000) to be demonstrated (see Figure 3, which is from Wild, Pfannkuch, Regan, & Horton, n.d.). The animations use memory-tracking of sample proportions in repeated samples and are built up slowly to start with, one sample at a time, and then shown together as a movie clip. The higher variability of the sample proportions in the 30% to 50% range compared to the lower variability of the sample proportions below 10% was raised and discussed with students.

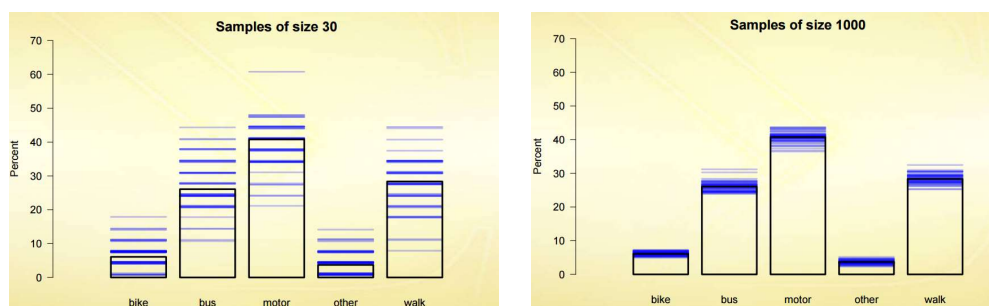


Figure 3. Animations showing a decrease in sampling variability with an increase in sample size

Conceptualising the margin of error for a single proportion A hands-on activity was designed to introduce the technique of bootstrapping to students. This simulation used the same categorical variable as Wild's (2009) animations, namely method of transport to school (Year-13 students also use bootstrap intervals for quantitative data). The population is an opaque bag of 605 data cards, each providing data on nine variables belonging to students at a fictitious Karekare College, but the student data on each card was drawn from actual students who participated in the Census-at-School survey (CensusAtSchoolNZ, n.d.). Each student took a sample of 25 cards and noted the percentage travelling to school by car. To remind students that, although imperfect, a random sample can still be considered to provide a reasonable picture of the population, Wild's (2009) rippled glass metaphor was evoked. Each student now took two re-samples from their sample, with the re-sample percentages placed on a line on a board to demonstrate the wide variability in poll percentages for samples of size 25. A sample of size 25 from Karekare College was pre-loaded into *iNZight* software (Wild, n.d.), with 36% of students in this particular sample travelling to school by car (Figure 4).

Although the software shows the sample proportion as 0.36, the verbal commentary with students rephrased this as 36% since percentages are the prevalent form of proportions in media reports. The re-sampling process was then repeated using the visualization tool. An important feature of the tool is the tracking of the animation which was switched off after running through the process a few times. This was followed by 20 and then 1000 re-samples (see Figure 5).

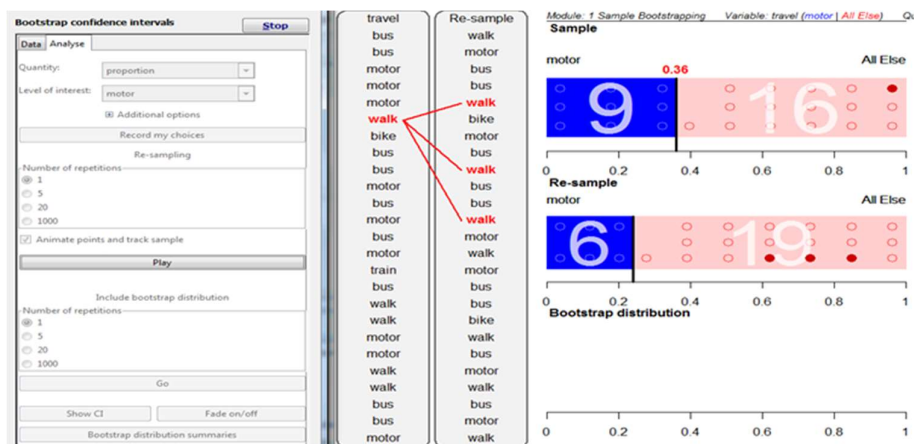


Figure 4. Re-sampling in the bootstrap confidence intervals module for proportions

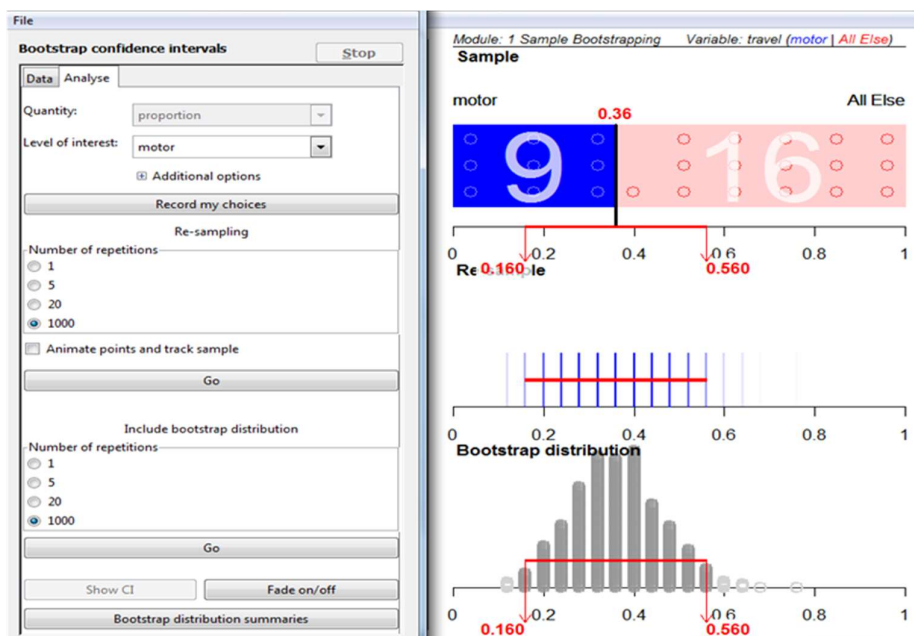


Figure 5. A bootstrap interval in the module for proportions

Students were then asked to suggest an interval within which most of the re-sample proportions lie. With a sample of size 25, this interval is very wide and the students suggested between 20% and 60%. The bootstrap interval, obtained using the central 95% of the re-sample proportions, was then added, showing as a red line on the scale below the bootstrap distribution, together with the numerical lower (0.160) and upper (0.560) limits. The interval also appears superimposed on the memory-tracking lines and below the original sample. It was hoped that the key words *confidence interval* would later prime recall of the visual image of the red line with its two numerical limits. Time constraints within this study restricted the use of software to a teacher-demonstration tool. However, when implementing the teaching sequence in classrooms, students would use the software themselves.

Conceptualising a 95% confidence interval for a single proportion We approximate classical confidence intervals by bootstrap intervals. Conceptually, the two methods are different but – in many cases – bootstrapping provides good approximations (for details, see Engel, 2010). The software offers a *coverage* module which allows for the development of conceptual understanding of 95% confidence (Wild, n.d.). The entire population of Karekare College (605 rows of data) was imported into the coverage module and the statistic of interest, the proportion of students who travel to school by car, displayed (see top right, Figure 6).

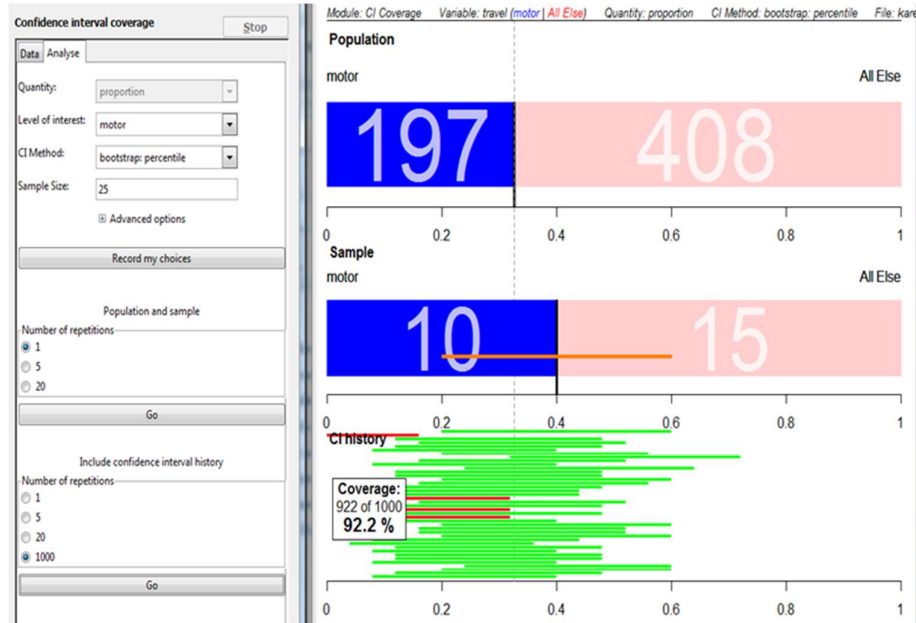


Figure 6. The coverage module for bootstrap intervals for a single proportion

A sample of size 25 was then selected and dropped down from the population to the rectangle below (middle right, Figure 6), showing those travelling by car as blue and the remainder as pink. The bootstrap confidence interval appears as a line superimposed on the sample box, with a copy travelling down to a scale at the bottom of the screen. The true population proportion travelling by car is shown as a vertical, dotted line running from the population at the top of the screen down to the scale at the bottom. A further 20 samples were then taken from the population of Karekare College, their confidence intervals generated, with those capturing the true population proportion shown as green lines, and those not capturing the true population proportion shown as red lines. The coverage percentage is shown in a white box at the side. When 1000 samples are taken, the software demonstrates that approximately 95% of the bootstrap confidence intervals capture the true population proportion which serves to validate the bootstrapping process. It was hoped that the key phrase *95% confidence* would later prime visual recall of the red and green lines in the coverage module image. The students' focus was then returned to the one sample that they had in reality, with the following argument used to develop an inferential statement:

We have demonstrated that one bootstrap interval calculated from that one sample (using 1000 resamples) has a 95% chance of capturing the true proportion of students

at Karekare College who travel to school by car. Hence we can infer, with 95% confidence, that this proportion is “somewhere between” 16% and 56%.

The margin of error is defined as half of the length of the bootstrap interval, which in this case, with a sample of only 25, is 20%. At this point, students were asked what would be needed in order to reduce the margin of error. Entering a sample of size 100 into the coverage module quickly demonstrates that using a sample four times bigger halves the width of the confidence interval and hence halves the margin of error. Reminding students that polls in the media usually have a sample size of at least 500, the bootstrap module was then repeated using a sample of 500, producing a much narrower bootstrap confidence interval and a margin of error of about 4.5%.

Introducing the rule-of-thumb for the margin of error for a single proportion The reasons for using the rule-of-thumb were twofold: first, unlike the formal statistical formula, this rule is easy to remember and calculate, thus providing an efficient way for data consumers to check a claim when reading a media report; and secondly, students do not have access to computers during their externally-assessed examination on statistical literacy. In NCEA, students can use either the rule-of-thumb or the formula

$$2 \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}},$$

both of which are taught in the course content. At this stage in the HLT, it was expected that students would have gained sufficient understanding of the margin of error to appreciate the $\frac{1}{\sqrt{n}}$ rule-of-thumb for estimating the margin of error for a poll percentage in the 30% to 70% range at a 95% level of confidence (Wild & Seber, 1999). Note that this approximation is optimal when the poll percentage is 50%:

$$2 \times \sqrt{\frac{0.5(1-0.5)}{n}} = \frac{1}{\sqrt{n}}.$$

It yields a too large number for large and small poll percentages, hence use of this rule is advisable only in the 30% to 70% range. Because $\frac{1}{\sqrt{25}} = 20\%$, $\frac{1}{\sqrt{100}} = 10\%$, and $\frac{1}{\sqrt{500}} = 4.5\%$, it can be demonstrated to students that this rule produces quick approximations consistent with the observed margins of error in the situations encountered so far in this learning sequence.

Being required to use this rule should reinforce the connection of sampling error to sample size. The media article (Appendix A) was re-visited to check this rule for the margin of error and through calculating a 95% confidence interval, checked the claim made in the article. The students were then asked to test four other simple claims made in media articles. Following a suggestion from the participating teachers, a procedural scaffold was developed (see Figure 7). This was provided for students in the initial stages of the teaching sequence, with the aim of encouraging students to develop their own procedural scaffolds as they moved to more advanced scenarios involving comparing poll percentages. More information on the visualization tools, simulation activities and teaching resources can be found in Rose (2012).

MoE =	Poll% =
95% CI:	
Interpret CI in words:	

Figure 7. Procedure for testing a simple claim

4. RESULTS

This section describes the students' reasoning and thinking as they experienced the different aspects of the learning trajectory with a focus on the margin of error and underlying concepts. Section 4.1 uses evidence from the eight pre-tests and four pre-interviews to describe the reasoning of students prior to the teaching intervention. Section 4.2 uses evidence gleaned from the two teaching sessions, the four post-tests and the four final interviews to describe the development of students' conceptual understanding of margin of error, while Section 4.3 compares the level of conceptual understanding from before to after the intervention using evidence from pre- and post-tests.

4.1. STATISTICAL LITERACY BEFORE TEACHING INTERVENTION

Question 1 in the pre-test (Appendix B) addressed students' ability to gain information from a piece of text, and their prior grasp of statistical concepts such as target population, random sampling, potential biases and non-sampling errors. All eight students were able to identify the funder and the purpose of the study. Four students showed confusion between the target population and the sample, describing the target population as though it were a sample: "people over 15 living in households." However, when interviewed, one of these students was able to clearly define the target group: "The group aimed at people over 15. Any citizen in New Zealand." All eight students had difficulty in articulating a reason for employing a random selection process. When probed further in the interviews, there appeared to be a tendency to fall back on stock phrases such as "a fair representation", with only a tenuous grasp of what that statement might mean.

I: What do you mean by a fair representation?

S1, S4:^a Different age-groups, gender, ethnicities.

S2: Different ethnic groups – an equal amount of people from each group.

S3: A variety of people, not just girls, different age-groups, different places in New Zealand.

I: What should a fair representation of the population of New Zealand have in it?

S1: Half of them male and half female.

I: The information states that the survey was representative of the population in terms of smoking status and ethnicity. What do you think representative means in this context?

S1: An equal number of people who smoke and don't smoke? I'm not really sure.

S4: I think you'd need equal numbers of people who smoke and don't smoke because you'd need the opinions of both.

I: Do you think the population of New Zealand has an equal proportion of smokers and non-smokers?

S1, S4: No.

I: So what might be a fair representation for New Zealand then?

S1: Maybe according to the numbers in the population. Convert that into a smaller size for the investigation.

S4: I don't know.

a. Note that interviews were carried out individually. Where students respond similarly, their responses have been combined.

It would seem that these students have understood that representative samples may comprise a variety of groups of people. However, they appear to believe that key characteristics such as gender, age, ethnicity, income, smoking status, etc., should occur in equal proportions in the sample. They do not appear to have grasped the key idea that such characteristics should be present in about the same proportion as they are in the population, and that this is best achieved by random sampling. Furthermore, when asked how the households in the survey might have been randomly selected, it would seem that their prior experiences of taking random samples were firmly embedded within the classroom setting, with responses such as “... select without knowing, like in class we get people’s names and draw out of a hat” (S2).

The pre-test interviews probed the students’ understanding of *making an inference*. Despite having studied this topic previously, none of the students were able to retrieve this formal statistical language. When asked what they understood about the statistical term making an inference, the students’ lack of confidence was evident from their replies which were often expressed as questions of the interviewer: “A hypothesis? Like when you make a prediction?” (S1); “Is it when you pose the question? No – wait – when you are making a statement?” (S3). However, they were able to use the more colloquial equivalent of *going back to the population*, and once it was established that making an inference was making a statement about the population, all four of the students who were interviewed quickly found an example of an inferential statement in the article in Question 2 (Appendix B).

Students were also asked in pre-interview if they were familiar with the term *sample-to-sample variability*. Their responses suggested that all four students understood the term although only one could readily articulate that the variability in the median for a sample of size 1000 would be much less than for a sample of size 30. All students noted that if random sampling of households was to be repeated, the poll percentages would differ slightly. Regardless of the fact that the concept of sample-to-sample variability appeared to be well-understood, none of the eight students in the pre-test could connect this to the term margin of error although two students recognized that $\pm 2.5\%$ should be added to 49.8% in Question 2. When probed further in the pre-interview, one student (S1) thought it might be “like the $\pm\%$ [measurement] error we add in physics when we do experiments,” but the other three students could not suggest anything.

Despite not yet having been taught about margin of error, and therefore not able to formally check claims made in media reports, two students demonstrated some statistical insight prior to the teaching intervention with regard to the following claim (Question 3f):

Popua claimed that agreement for the statement “Cigarettes and tobacco should not be sold in New Zealand in 10 years’ time” was 11.8% higher among Pacific Islanders than among Maori.

Student S1 recognized that different sample sizes for Maori and Pacific Islanders would affect the validity of the claim while student S2 went further by stating: “The number of Pacific Islanders is very small. I think you’d need to consider both the percentages [pointing to the table] and the sample sizes.”

Reflecting on student responses to the pre-tests and pre-interviews, it appeared that all students would benefit from literacy support provided through the HLT. It was envisaged that being taught pre-reading and while-reading strategies would assist students when grappling with unfamiliar texts and would broaden their statistical horizons. Although students could be expected to have some statistical knowledge prior to the teaching intervention, concepts such as the effects of sample size, random sampling, defining variables, interpreting tabular information, and familiarity with formal statistical language, would require further development.

4.2. STUDENTS' UNDERSTANDING OF MARGIN OF ERROR AND 95% CONFIDENCE INTERVAL

In this section, classroom interactions, post-tests and post-interview responses are used to illustrate the development of students' conceptual understanding of margin of error. In particular, analysis will focus on conceptualizing the margin of error; conceptualizing a 95% confidence interval; and using rules-of-thumb to test claims.

Conceptualising the margin of error The chosen media article on military exercises (Appendix A) provided a satisfactory base for the development of the concept of margin of error with the added advantage of demonstrating to students how to use information within a text in conjunction with prior knowledge and experience to make sense of unfamiliar language. The students were sufficiently familiar with the term ANZAC and New Zealand's non-nuclear policy to quickly make sense of ANZUS rift and, as the second author read the text aloud, she asked questions to check student comprehension. The discrepancy in sample size, 850 in the graphic but 750 in the text, was fortuitous as it enabled the students to use their newly-acquired rule-of-thumb to resolve the discrepancy.

As anticipated, the students initially wanted a sample of size 30 in the hands-on activity (see Section 3.2), stating "we always use 30". However, as conjectured, the animations appeared to provide a vital link in enabling the students to move from their prior understanding, constructed from experience with measured data, to a realisation that categorical data had much higher sampling variability with small samples and therefore considerably higher sample sizes were needed in polls and surveys. For one student (S3), it was the animations that she most vividly recalled in her post-interview. While her post-test response to Question 2a demonstrated her apparent awareness of the connection between margin of error and sample size (Figure 8 left), when asked to draw the image which came to mind when the researcher said the term confidence interval, it was Wild's (2009) animation that S3 drew (Figure 8 right), stating "those bars that were going up and down." Thus the animations appeared to provide a vital first step in her understanding.

(a) Explain to someone who has not done statistics how to interpret the margin of error.

The bigger the sample size, the smaller the margin of error. The margin of error is given to show the variability. If another sample was to be taken the percentages would be within 2.5% of the original sample taken.



Figure 8. Left. Student S3's response to Question 2a (Appendix B).
Right. Student S3 recalls animations for sampling variability for categorical data

The hands-on activity also appeared to be helpful in improving students' conceptual understanding. For example, S5, in her post-interview about her understanding of the phrase "representative of the population in terms of ethnic origin," described the sample as "an opaque view of each of the different ethnic origins," possibly recalling the placing of the sample in an opaque bag in the hands-on activity as a concrete representation of Wild's (2009) rippled glass metaphor. With further probing, she clarified her view of the ethnic distribution of New Zealand as "a small percentage of Maori and Pasifika with lots of Asian and European", indicating that her understanding did now encompass some proportionality in the representation of groups. Student S3 also recalled the taking of cards out of the bag and "being sure to replace it" when asked to describe a resample.

The following four responses to Question 2 (Appendix B) in the post-tests illustrate the students' conceptual understanding of the statistical term *margin of error* at the end of the teaching intervention:

- S1: The margin of error is an indication of how much sampling error there is. The actual % is likely fall within the % stated in the paper \pm the margin of error.
- S2: Margin of error is needed when we refer back to the population. There are sample-to-sample variability. Margin of error shows the percentage ranges of a particular group or groups, showing that if another sample were to be carry out again, the percentages of a particular group will more than likely fall between the margin of error of the [original] sample percentage. Margin of error is half of the confidence interval.
- S3: It explains how big a difference there could be if another sample was taken.
- S5: Since the people should get the idea the poll percentage may vary as it is not an exact representation of the whole population.

Since the students had no prior knowledge of the concept of margin of error, we conjecture that the views outlined above were gained during the teaching intervention. The use of the terms *sampling error*, *sample-to-sample variability* and *poll percentage* in the student responses suggests that they were beginning to develop notions of margin of error and were feeling more comfortable using formal statistical language. Notably, these four students used the rule-of-thumb to verify the reported margin of error in the article.

Conceptualising a 95% confidence interval Although student S3 seemed to gain some understanding of margin of error and its relationship with sample size, she was not able to consistently use this information to calculate a confidence interval. However, she was able to interpret a given confidence interval in words. Throughout her post-test, S3 multiplied the poll percentage by the margin of error, instead of adding. She did not readily recall a confidence interval as a line, although when shown Figure 5 in her post-interview and asked to point to the confidence interval, she pointed to the red line superimposed on the blue bars in the middle section. When asked what she would need to do to expand her answer to Question 2a, (Figure 8), S3 responded that she would need to calculate the confidence interval. When asked to carry out the calculation, she became confused. The interviewer conjectured that drawing a line to represent the confidence interval would help S3:

- I: Remember the picture we have just looked at? [pointing again to Figure 5]. Did we agree that a confidence interval is this line in the middle? [S3 nods.] OK, so I'm drawing a line here [on a sheet of paper] for you. Where would you put the 49.8%?
- S3: Um, just here, below half-way [pointing underneath the middle].
- I: OK, in the middle [I writes in the 49.8%]. So where is the 2.5%, the margin of error?
- S3: It would be 2.5% this way and 2.5% that way. [S3 traces her finger along the line from the middle in each direction.]
- I: So now do you think you can calculate the confidence interval?

S3 then correctly subtracted and added the 2.5% to the 49.8% and correctly interpreted the interval in words, saying: "with 95% confidence we can claim that people

that agree that cigarettes should no longer be sold in New Zealand in ten years' time would be somewhere between 47.3% and 52.3%". Later in the interview, following a prompt to draw a diagram, S3 corrected her post-test answer to a subsequent question and correctly interpreted that confidence interval in words. This finding suggests that drawing a diagram when calculating a confidence interval can be helpful.

S3's understanding of the expression *with 95% confidence* was further probed. Although she could not recall the coverage module well enough to draw it herself, S3 correctly interpreted Figure 6, stating that the dotted line represented the true population percentage and that most of the confidence intervals would capture it, while a few would not. However, this was not yet easily accessible to S3, requiring some thinking time for her to successfully retrieve the information. The other three students (S1, S2 and S5), however, more easily recalled the coverage module in their post-test interviews. S2 drew a diagram showing the vertical line with the true population proportion and horizontal lines both crossing and not crossing the vertical line (Figure 9).



Figure 9. Student S2 recalls an image of the coverage module in a post-test interview

S1 and S5 drew similar diagrams with all three students mentioning the green confidence intervals capturing the true percentage most of the time, and the red confidence intervals sometimes missing it. The coverage module seemed to serve as an important visualization tool that assisted these students to conceptualise a 95% confidence interval, in line with the conjecture of the HLT proposed for this study.

4.3. COMPARING PRE- AND POST-TEST REGARDING MARGIN OF ERROR

In this section, the performance of the students on the pre-test and the post-test is assessed overall for the four students who completed both tests.

The assessment schedule was modelled on that used by the New Zealand Qualification Authority (NZQA) in 2011 which was, in turn, based on SOLO taxonomy (Biggs & Collis, 1982). In the schedule, assessment items graded *u* demonstrate understanding at a multi-structural level (Achievement), those graded *r* at a relational level (Merit), and those graded *t* at an extended abstract level (Excellence). For the margin of error topic, there are eight possible levels of performance from low Not Achieved (N1) through to high Excellence (E8). In order to gain Achievement at a particular level, at least half of the items must be attained at that level (see Appendix C). Note that the schedule includes items that relate to comparison of poll percentages, not discussed in this paper (see Rose, 2013, for further details). Each student was allocated a score which was used to assess the overall level of statistical reasoning attained regarding the margin of error in polls and surveys.

It is not surprising that all four students demonstrated an increase in performance from the pre-test to the post-test. Three of the students [S1, S2 and S5] were able to use formal statistical language in their post-test responses, could extract mathematical information from tables and used it to test claims made in media reports, and demonstrated an appreciation for the need for a margin of error regarding such claims.

However, we believe that these students were unable to complete one SOLO-like progression as conjectured by Watson (2006), since they were unable to demonstrate the ability to critically evaluate a media report with statistical insight without the prompts afforded by the scaffolding in the post-test and further questions from the interviewer in their post-interviews.

Table 1. Assessment of student pre-test and post-test using the schedule in Appendix C

Student	Pre-test		Post-test		Change
	Level	Mark	Level	Mark	
S1	$2 \times u = N2$	2	$4 \times u + 3 \times r = M5$	5	+3
S2	$1 \times u = N1$	1	$1 \times u + 2 \times r + 2 \times t = E7$	7	+6
S3	$0 = N0$	0	$2 \times u = N2$	2	+2
S5	$1 \times u = N2$	2	$3 \times u + 2 \times r + 1 \times t = M5$	5	+3

5. DISCUSSION

The main aim of this paper was to investigate what elements of a specifically designed statistical literacy hypothetical learning trajectory could facilitate students' conceptual access to the ideas underpinning margin of error.

The use of media reports appeared to provoke both an initial interest and a purpose for the conceptual development of margin of error. However, the learning trajectory had to extend students' existing knowledge base from their experience of the relatively small sampling variability of quantitative data and the use of relatively small sample sizes to the much wider sampling variability of categorical data and hence the need for much larger sample sizes.

In accordance with Rossman (2008), animations and simulations using visualization software, imitated in a concrete hands-on activity prior to using the software, appeared to be valuable tools in aiding students' conceptual development of margin of error and 95% confidence intervals. Although the need for a rule-of-thumb for calculating a margin of error seemed to be accepted by the students, they wanted to be convinced of its veracity. The *iNZight* modules provided a seemingly adequate demonstration of the $\frac{1}{\sqrt{n}}$ rule for a single poll percentage. However, it would seem that encouraging students to draw a confidence interval as a line, with the sample percentage in the middle, will further assist their understanding. In terms of checking claims, procedural scaffolding to reduce cognitive overload was also an important requirement (Van Merriënboer & Sweller, 2005).

Consistent with dual coding theory proposed by Clark and Paivio (1991), Wild's (2009) rippled glass metaphor for sampling, imitated in the use of opaque bags in the concrete hands-on activities, seemed powerful in helping students accept the bootstrapping procedure for approximating a confidence interval. Similarly, the term *confidence interval* and/or the phrase *with 95% confidence* became appropriately linked to a visual image in the coverage module, with three of the four participants readily recalling this when asked to conceptualise a 95% confidence interval in their post-interview. Thus these Year-12 students were able to make reasonable progress in the area of margin of error.

Having implemented the teaching intervention, several issues in students' reasoning processes have arisen. The first issue is that the Year-12 student participants who took part in this teaching intervention appeared to have only a superficial knowledge of the

history and geography of New Zealand and current hot topics in the media. They generally did not read newspapers or follow the news in other media. They therefore lacked the broad contextual knowledge required to critically evaluate statistically-based reports. The study of media reports may therefore have an additional role to fulfil in a statistical literacy course: that of prompting students to broaden their general knowledge.

A second issue, a misconception regarding representativeness and the nature of random samples, may be connected to the first issue. This misconception involves the suggestion on the part of the students that key characteristics (age, income, ethnicity, regions of residence, etc.) should occur in equal proportions in the sample rather than in the proportions in which they occur in the New Zealand population. This misconception may have its roots in the fact that random sample items have an equal probability of selection. We conjecture that this misconception is further exacerbated by the lack of knowledge on the part of the students regarding the distribution of these key characteristics in the New Zealand population. The predominance of comparative investigations in earlier school years, using approximately equal sample sizes for the compared groups, may also be a factor aggravating this misconception. Thus an additional activity will need to be devised which enables students to explore the distributions of key characteristics both in the wider New Zealand population and in the population of the students' locality. Furthermore, since these students had difficulty in suggesting how a random sample might be selected, opportunities to learn about a variety of real-world random selection methods should be sought so that students' views can extend beyond their classroom experiences.

There are several limitations to this study. Firstly, the participating teachers who contributed to the HLT were volunteers who may not be typical of future teachers. Thus the improvements to the HLT on the basis of first cycle of the study may not be definitive. Secondly, the teaching intervention, implemented with a very small group of female students, was conducted outside the students' normal teaching programme with only half of them completing both teaching sessions. Additionally, the students were all from the same high socio-economic girls' school, therefore not typical of the students who might undertake a Year-13 course on statistical literacy, and therefore the findings can only be considered to be an indication of how Year-13 students might perform. Thirdly, time constraints of the study limited the use of the software to that of a teacher-demonstration tool, with no opportunity for the students to consolidate their learning through personal interaction with the software.

Statistical literacy has become an essential skill needed to function as an effective citizen in today's world (Gal, 2002; Rumsey, 2002b; Wallman, 1993). However, statistical illiteracy is widespread (Gigerenzer et al., 2007; Goldin, 2010), requiring all data consumers to be able to adopt a critical stance and have the skills needed to assess the validity of any statistical information they encounter. As this small study seems to demonstrate, the ability to critically evaluate statistical information requires a critical stance, literacy skills, mathematical knowledge, contextual knowledge, and a reasonable level of statistical understanding (Gal, 2002; Watson, 2006). Also these components are not generally acquired during general education and must be specifically taught (Noll, 2007; Pryor, 2001; Watson, 2006). To date, there have only been a few studies involving attempts to teach statistical literacy and report on students' learning outcomes (e.g., Budgett & Pfannkuch, 2010; Merriman, 2005; Nicholson, Ridgway, & McCusker, 2010), much of it small-scale. However, the components of the learning trajectory for building statistical literacy specifically designed in this study may have the potential to facilitate students' conceptual access to ideas underpinning margin of error.

ACKNOWLEDGEMENTS

We thank Maxine Pfannkuch for her helpful comments on this paper. We also thank the anonymous reviewers of this paper for their constructive suggestions and comments.

REFERENCES

- Best, J. (2004). *More damned lies and statistics: How statistics confuse public issues*. Berkeley, CA: University of California Press.
- Biggeri, L. & Zuliani, A. (1999). The dissemination of statistical literacy among citizens and public administration directors. *Bulletin of the International Statistical Institute 52nd session*. Voorburg, The Netherlands: International Statistical Institute. [Online: www.stat.fi/isi99/proceedings.html]
- Biggs, J. & Collis, K. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York, NY: Academic Press.
- Blastland, M. & Dilnot, A. (2010). *The numbers game: The commonsense guide to understanding numbers in the news, in politics, and in life*. London: Gotham.
- Budgett, S. & Pfannkuch, M. (2010). Using media reports to promote statistical literacy for non-quantitative majors. In C. Reading (Ed.), *Proceedings of the Eighth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute.
- Clark, J. M. & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, 3(3), 149–210.
- CensusAtSchoolNZ (n.d.). *Where NZ teachers come for information and support for their teaching of statistics*. [Online: new.censusatschool.org.nz/]
- Cobb, P. & Gravemeijer, K. (2008). Experimenting to support and understand learning processes. In A. E. Kelly, R. A. Lesh, & J. Y. Baek (Eds.), *Handbook of design research methods in education*. (pp. 68–95). New York: Routledge.
- Engel, J. (2010). On teaching bootstrap confidence intervals. In C. Reading (Ed.), *Proceedings of the Eighth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute.
- Everitt, B. S. (1999). *Chance rules: An informal guide to probability, risk and statistics*. New York: Springer.
- Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., & Perry, M. (2007). *Guidelines for assessment and instruction in statistics education (GAISE)*. Report: a pre-K-12 curriculum framework. Alexandria, VA: American Statistical Association. [Online: www.amstat.org/education/gaise/GAISEPreK-12_Full.pdf]
- Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities (with discussion and a rejoinder by the author). *International Statistical Review*, 70(1), 1–51.
- Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest*, 8(2), 53–96.
- Goldin, R. F. (2010). Spinning heads and spinning news: The American media's gap in quantitative reasoning skills. In C. Reading (Ed.), *Proceedings of the Eighth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute.
- Merriman, L. (2007). *Using media reports to develop statistical literacy in year 10 students* (Unpublished master's thesis). The University of Auckland, Auckland, New Zealand.
- Ministry of Education. (2007). *The New Zealand curriculum for English-medium teaching and learning in years 1–13*. Wellington, New Zealand: Learning Media. [Online: nzcurriculum.tki.org.nz/The-New-Zealand-Curriculum]

- Moore, D. S. (1998). Statistics among the liberal arts. *Journal of the American Statistical Association*, 93(444), 1253–1259.
- Nicholson, J., Ridgway, J., & McCusker, S. (2010). Luring non-quantitative majors into advanced statistical reasoning (and luring statistics educators into real statistics). In C. Reading (Ed.), *Proceedings of the Eighth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute.
- Noll, J. A. (2007). *Graduate teaching assistants' statistical knowledge for teaching* (Doctoral dissertation). Portland, OR: Portland State University, USA.
[Online: iase-web.org/Publications.php?p=Dissertations]
- NZ Herald (2012). *Opinion divided on NZ-US exercises*. Article 14 May 2012.
- Pryor, H. (2001). Assessment of the statistical literacy ability of some tertiary students using media reports (Unpublished master's thesis). The University of Auckland, Auckland, New Zealand.
- Ridgway, J., Nicholson, J., & McCusker, S. (2007). Teaching statistics – despite its applications. *Teaching Statistics*, 29(2), 44–48.
- Ridgway, R. A. & Ridgway, J. (2010). Crimes against statistical inference: Forcing teachers to be accessories after the (absence of) fact. *Online Educational Research Journal*, 11, 1–11. [Online: www.oerj.org/View?action=viewPDF&paper=27]
- Rose, D. (2012). Margin of Error and testing claims in the media. *CensusAtSchoolNZ*.
[Online: new.censusatschool.org.nz/resource/margin-of-error-and-testing-claims-in-the-media/]
- Rose, D. (2013). Developing statistical literacy in year 13 students (Unpublished Master's thesis). The University of Auckland, Auckland, New Zealand. [Online: researchspace.auckland.ac.nz/bitstream/handle/2292/20042/whole.pdf?sequence=2]
- Rossmann, A. (2008). Reasoning about informal inference: One statistician's view. *Statistics Education Research Journal*, 7(2), 5–19.
[Online: iase-web.org/Publications.php?p=SERJ_issues]
- Rumsey, D. J. (2002a). Discussion: Statistical literacy: Implications for teaching, research, and practice. *International Statistical Review*, 70(1), 32–36.
- Rumsey, D. J. (2002b). Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education*, 10(3).
[Online: ww2.amstat.org/publications/jse/jse_archive.htm]
- Schild, M. (2002). Three kinds of statistical literacy: What should we teach? In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute.
- Schild, M. (2004). Statistical literacy and liberal education at Augsburg College. *Peer Review*, 6(4), 16–18. [Online: www.statlit.org/pdf/2004SchildAACU.pdf]
- Smith, A. F. (2004). *Making mathematics count*. London: The Stationary Office
[Online: www.mathsinquiry.org.uk/report/MathsInquiryFinalReport.pdf]
- Van Merriënboer, J. J. & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, 17(2), 147–178.
- Wallman, K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association*, 88(421), 1–8.
- Watson, J. M. (2002). Discussion: Statistical literacy before adulthood. *International Statistical Review*, 70(1), 26–30.
- Watson, J. M. (2006). *Statistical literacy at school: Growth and goals*. Mahwah, NJ: Lawrence Erlbaum.
- Wild, C. J. (2009, May). Early statistical inferences: “The eyes have it”. Keynote Talk, United States Conference on Teaching Statistics. Columbus, Ohio.
[Online: www.stat.auckland.ac.nz/~wild/09.USCOTSTalk.html]

Wild, C. (n.d.). *VIT: Visual Inference Tools*. The University of Auckland.
 [Online: www.stat.auckland.ac.nz/~wild/VIT/index.html]
 Wild, C. & iNZight team (n.d.). *iNZight*. Software. The University of Auckland.
 [Online: www.stat.auckland.ac.nz/~wild/iNZight/index.php]
 Wild, C. J. & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry (with discussion). *International Statistical Review*, 67(3), 223–265.
 [Online: iase-web.org/documents/intstatreview/99.Wild.Pfannkuch.pdf]
 Wild, C., Pfannkuch, M., Regan, M., & Horton, N. (2011). Towards more accessible conceptions of statistical inference (with discussion and a reply from the authors). *Journal of the Royal Statistical Society, A 174*, 247–295. [Online: doi.org/10.1111/j.1467-985X.2010.00678.x]
 Wild, C., Pfannkuch, M., Regan, M., & Horton, N. (n.d.). *Animations of sampling variation*. WPRH. [Online: www.stat.auckland.ac.nz/~wild/WPRH/]
 Wild, C. J. & Seber, G. A. (1999). *Chance encounters: A first course in data analysis and inference*. New York, NY: Wiley.

STEPHANIE BUDGETT
 Department of Statistics, The University of Auckland
 Private Bag 92019, Auckland,
 New Zealand

APPENDIX A: ACTIVITIES FOR ASSESSING MARGIN OF ERROR

Media article and activities for reading an unfamiliar context introducing margin of error.

Opinion divided on NZ-US exercises

Military exercises between New Zealand and the United States are being stepped up but the public is fairly divided over them.

New Zealand Army engineers will train in the United States with the Marine Corps next month in an exercise dubbed Galvanic Kiwi, the first joint exercise on American soil since the Anzus rift over New Zealand anti-nuclear policy in 1984.

It will follow closely on the heels of the Alam Halfa exercise in New Zealand which involves 76 US military personnel, the first combat exercises since the rift.

A Herald-DigiPoll of 750 voters conducted shortly after the announcement of the Alam Halfa exercise suggests more people approve of the resumption of exercises, but not overwhelmingly.

Asked, "Do you approve or disapprove of the United States resuming military exercises in New Zealand," 47.6 per cent approved, 44 per cent disapproved and 8.4 per cent didn't know or wouldn't say.

The US exercise will involve 36 New Zealand defence personnel. It will be a reciprocal platoon exchange with 50 Marines who will visit New Zealand next month with a Marine band to mark the 70th anniversary of the arrival of the Americans in World War II to aid New Zealand's defence.

Defence Minister Jonathan Coleman said he was a little surprised and thought support for the exercises might be higher but respondents might have thought it involved ship visits.

"There'll always be a hard core who will disapprove." — Audrey Young

“Opinion divided on NZ-US exercises” (NZ Herald, 2012)

1. Read the article and then write a “gist statement” (3 sentences) summarising what the article is about.

2. Margin of Error:

For poll %s of about 50% (between 30% and 70%), margin of error \cong at a confidence level.

For poll %s **below 30% and above 70%**, the margin of error is

95% of the time, the 95% confidence interval the true percentage in the population.

We can say, with 95% confidence from a sample of 500, that the % of NZ children who travel to school by car is **somewhere between and**

3. Test the claim in the article that “opinion is fairly divided”

Margin of error: (a) rule-of-thumb = (b) reported MoE =

Sample % who support resumption of exercises =

95% confidence interval: Meaning Judgment



APPENDIX B: STUDENT PRE- AND POST-TEST

The tasks have been adapted from STATS 150 course, Dep. Statistics, U. of Auckland.

All questions in this test relate to a Health and Lifestyle survey conducted in 2008.

Question 1.

The Health Sponsorship Council commissioned the National Research Bureau to conduct a Health and Lifestyle Survey from early May to the end of July in 2008. The survey involved face-to-face interviews with 1608 people in their homes. Households from throughout New Zealand were randomly selected. From each household, the person aged 15 years and over who had the next birthday was asked to take part in the survey. Data was post-weighted to ensure it was representative of the population in terms of smoking status and ethnic origin.

In the questionnaire participants were given the following 3 statements about cigarettes:

1. Cigarettes and tobacco should not be sold in New Zealand in 10 years' time.
2. Tobacco companies should not be allowed to promote cigarettes by having different brand names and packaging.
3. The number of places selling cigarettes and tobacco should be reduced to make them less easily available.

For each statement they were asked to respond: Strongly Agree, Agree, Neither Agree nor Disagree, Disagree or Strongly Disagree.

Read the above description of the Health and Lifestyle survey then answer the following questions:

- (a) Who funded this survey?
- (b) What was the main purpose of the survey?
- (c) State the target population for this survey?
- (d) In your own words,
 - (i) Briefly describe how people were selected to take part in this survey.
 - (ii) Why do you think this selection method was used?
- (e) Identify and discuss two concerns that you may have with this survey.

Question 2.

Kiwis want cigarettes banned by 2020

4:00 AM Friday Jan 29, 2010

Half the nation, including smokers, support completely banning cigarettes within 10 years, a study has found.

The 2008 Health and Lifestyle Survey compiled nationwide interviews from the Health Sponsorship Council of 1608 people, including 422 smokers, and has just been published in the NZ Medical Journal.



Photo / Kenny Rodger

It found 49.8 per cent of people agreed cigarettes should no longer be sold in New Zealand in 10 years, 30.3 per cent disagreed and 19.9 per cent neither agreed nor disagreed. Of the smokers surveyed, 26.2 per cent agreed and 55.3 per cent disagreed. The study also showed public support for plain, unbranded cigarette packets and fewer tobacco retailers.

Pacific Islanders, in particular, showed strong support for the measures.

It is reported in the NZ Herald article above that 49.8% of people agreed that cigarettes should no longer be sold in New Zealand in 10 years' time. The article stated that the margin of error was 2.5%.

- (a) Explain to someone who has not done statistics how to interpret the margin of error.
- (a) Why is a margin of error needed when reporting poll percentages?
 - (b) Use the information given in the NZ Herald article to show how the margin of error of 2.5% can be calculated.

Question 3.

Consider the data shown in Table 1 below on *Support for the end of tobacco sales within 10 years*. Note that for reporting purposes the Strongly Agree and Agree responses were combined into one category, **Agree**, the Strongly Disagree and Disagree responses into one category, **Disagree**.

Table 1: Support for the end of tobacco sales within 10 years: “Cigarettes and tobaccos should not be sold in New Zealand in 10 years’ time.

	Total	Maori	Pacific	Asian	Other	Never	Current	Past
Agree	49.8%	47.9%	59.7%	62.1%	48.2%	60.1%	26.2%	52.0%
Neither agree nor disagree	19.9%	17.3%	20.2%	15.9%	20.7%	18.0%	18.4%	23.1%
Disagree	30.3%	34.9%	19.7%	22.0%	31.1%	21.8%	55.3%	24.9%
Number surveyed (weighted)	1608	241	112	129	1126	616	378	614

The survey questionnaire needed clear, unambiguous definitions to ensure that participants’ smoking status was classified in the same way.

- (a) If you were developing classifications for smoking status, state two issues you need to think about.
- (b) Write your own definition of a *current smoker*.

To answer the following questions, use Table 1. You may need to do some calculations.

- (c) (i) What percentage of current smokers agreed with the statement “Cigarettes and tobacco should not be sold in New Zealand in 10 years’ time”?
- (ii) What percentage of participants was classified as either Maori or Pacific Islanders?
- (iii) How many Pacific Islanders agreed with the statement “Cigarettes and tobacco should not be sold in New Zealand in 10 years’ time”?

- (d) The NZ Herald article claimed that “Pacific Islanders, in particular, showed strong support for the measures.”

Alison argued that the newspaper was correct because 59.7% of Pacific Islanders surveyed agreed, and the reported margin of error was 2.5%

Comment on Alison’s argument.

You may wish to support your reasoning with appropriate calculations (see Table 1).

- (e) Hekia claimed that Maoris were more likely to agree than disagree with the statement “Cigarettes and tobacco should not be sold in New Zealand in 10 years’ time.”

Explain, with calculations, whether this claim is justified. (See Table 1).

- (f) Popua claimed that agreement for the statement “Cigarettes and tobacco should not be sold in New Zealand in 10 years’ time” was 11.8% higher among Pacific Islanders than among Maori.

Explain, with calculations, whether this claim is justified (see Table 1).

APPENDIX C: SOLO-LIKE SCHEDULE FOR ASSESSMENT

The following schedule has been used to assess student understanding of margin of error.

<i>Sampling error and 95% confidence intervals for proportions</i>				
Task	Evidence	Achievement (<i>u</i>)	Merit (<i>r</i>)	Excellence (<i>t</i>)
2 (a)	With 95% confidence, we can infer that the % of NZs aged 15+ who agree that cigarettes and tobacco should not be sold in 10 years' time is somewhere between 47.3% and 52.3%.	Response indicates that actual percentage who agree is likely to be somewhere between 47.3% and 52.3%.	Full interpretation including a reference to the target population and 95% confidence.	
(b)	A poll % is an estimate of the true population %. Taking another sample gives a different estimate. MoE gives a range of plausible values for the true %.		Comment indicates that MoE gives a range of plausible values for the true population %.	
(c)	Using the $1/\sqrt{n}$ rule-of-thumb: $1/\sqrt{1608} = 0.025 = 2.5\%$.	$1/\sqrt{1608} = 0.025 = 2.5\%$		
3 (c) (i) (ii) (iii)	26.2% 22.0% 23 people.	At least two of 3 (c) correct.		
(d)	Both Asians and Pacific Islanders (PI) show strong support for the proposal that cigarettes and tobacco should not be sold in 10 years' time.	Response acknowledges poll % for Asians is higher than for PI OR correctly applies rule-of-thumb $1/\sqrt{112}$ to note support from PI is between 50.3% and 69.1%.	95% CI for PI and comments support of PI is strong: lower limit is over 50% OR explains with some evidence that wording of claim should be amended, e.g., remove "in particular" or change PIs to "PIs and Asians" in particular".	Either calculates CI for difference OR calculates CI for both Asians and PI and explains both lower limits over 50% so claim should read "in particular, both PI and Asians show strong support".
(e)		Calculates MoE for difference using correct rule-of-thumb $= 2 \times (1/\sqrt{241}) = 12.8\%$	Calculates 95% CI for difference $= [0.2, 25.8]$ and correctly explains claim can be supported since lower limit is positive	We can infer with 95% confidence that the difference in % of Maoris who agree compared to the % who disagree is between 0.2 and 25.8 percentage points .
(f)		Calculates MoE for difference using $1.5 \times \text{average MoE} = (6.4 + 9.4)/2 = 7.9\%$.	Calculates 95% CI for difference $= [-0.05, 23.8]$ and correctly explains claim cannot be supported since lower limit is negative.	We can infer with 95% confidence that the % support from PI could be up to 0.05 percentage points less and up to 23.8 percentage points more than the % support from Maoris.
Sufficiency requirements	N1: $1 \times u$ N2: $2 \times u$	A3: $3 \times u$ or $2 \times u + 1 \times r$ A4: $4 \times u$ or $3 \times u + 1 \times r$ or $2 \times u + 2 \times r$	M5: $3 \times r$ or $2 \times r + 1 \times t$ M6: $4 \times r$	E7: $3 \times r + 1 \times t$ or $2 \times t$ E8: $3 \times t$