Two Test Items to Explore High School Students' Beliefs of Sample Size when Sampling from Large Populations

Anthony Bill
University of Tasmania
<afbill@utas.edu.au>

Sally Henderson
Department of Education, Tasmania
< sally.henderson@education.tas.gov.au >

John Penman
Department of Education, Tasmania
< john.penman@education.tas.gov.au>

Two test items that examined high school students' beliefs of sample size for large populations using the context of opinion polls conducted prior to national and state elections were developed. A trial of the two items with 21 male and 33 female Year 9 students examined their naïve understanding of sample size: over half of students chose a sample size of "10% of the population", and a quarter chose a sample size of 15,000 – both approaches grossly exceeding the accepted sample size.

Formal consideration of sample size when designing a survey is presently largely excluded from the high school curriculum; it is a complex topic that high school students may not have the conceptual development to support. When a survey is conducted in schools, the most convenient population is the students present in the class: the survey is a census, and the sample size is the number of students present. Sample size is considered only when a survey is conducted beyond the confines of the classroom, and the practical difficulties for a teacher may discourage this type of investigation. When the opportunity for designing a survey – as distinct from a census – arises, students naturally ask the question "how many should be asked?" This suggests not only awareness that a representative sample imperfectly reflects the population, but also a desire to obtain as accurate a result as possible.

Theoretical Background

Statisticians and educators have observed the widespread misunderstanding of sample size amongst students, the general public, and within the media (e.g., Fielding, 1997; Smith, 2004). As a statistician, Fielding noted that inadequately trained investigators were often preoccupied with sample size as a fraction of the population, rather than the absolute sample size. In a study of college students, Smith observed similarly that most untutored students would use a sample size based on the size of the population, such as "10% of the population", and that many students found the notion that a larger population did not require a larger sample as counter-intuitive. Sample size was formerly part of the senior high school curriculum (e.g., Harding, 1992), and more recently studied at college level using computer based re-sampling techniques (e.g., Smith, 2004). Within schools the importance of representative and random sampling is emphasised, but the curriculum is curiously silent on the complementary topic of explicitly quantifying sample size (Department of Education, Tasmania, 2008). The explanation for this apparent omission may be quite simple: in a crowded curriculum this topic is displaced by what are considered higher priority learning goals, and the statistics literature provides no accessible alternative for high school students and school teachers than the crude "10% of the population" rule possibly first acquired in upper primary. Watson's (2006) extensive work with primary and middle school students examining statistical literacy – a cohort that includes the Year 9 students in the proposed study – considered sample size, but the work focused on part/whole concepts and sample representativeness rather than the explicit determination of sample size. Students encounter large populations through the media, for

L. Sparrow, B. Kissane, & C. Hurst (Eds.), *Shaping the future of mathematics education: Proceedings of the 33rd annual conference of the Mathematics Education Research Group of Australasia.* Fremantle: MERGA.

example political opinion polls, and the *CensusAtSchool* program (Australian Bureau of Statistics, 2009) now provides large data sets where analysis is feasible only through sampling of the data. The literature does not formally define a large population, but the populations in the test items are arguably large. The populations are also finite: "10% of the population" has no application with the *infinite* populations that students do encounter at school, such as die and coin systems. Formally quantifying students' beliefs using the two proposed test items, and with more sophisticated sample size models potentially accessible to high school students (to be published subsequently) as students mature mathematically, suggests that greater consideration of sample size may be warranted.

Method

The sample consisted of two Year 9 extended mathematics classes in two single-gender government high schools in an Australian capital city. The first author taught the classroom component of the research as part of research into the use of *Fathom*™ in high schools, and the second and third authors acted as colleague teachers. A theme of the research was an exploration of sample size, and the two test items presented below were used to examine students' naïve beliefs of sample size when sampling from large populations.

Both groups were defined as extended mathematics classes, but the students had self-selected to enrol in the course and presented with a range of abilities. While 21 male students and 35 female students participated, two female students were not fully competent English speakers and, given the language requirements and the contextual nature of the tasks, their responses were not included. The students were either 14 or 15 years old.

Description of the Two Test Items: Task 1 and Task 2

The two test items were presented as multiple-choice companion tasks: Task 1 and Task 2, using the context of opinion polls conducted prior to national and state elections. Both tasks examined sample size in large populations, but Task 1 National Election considered a population ten times that of Task 2 State Election. The number of voters is plausible, but not accurate. The items were designed to be familiar and experiential: the national election and the opinion polling were conducted seven months prior to the study; the students were not eligible to vote at that election, but would vote at the subsequent election. The items were presented under traditional examination conditions. The two items posed the same questions and offered the same responses, but Task 2 also included a response that related population and sample size. Both test items asked students to first, choose the sample size, and second, provide an explanation for that choice. These will now be discussed in turn.

When choosing a sample size, students were provided with four alternative responses that may be categorised as either a fixed percentage of the population, or a numerical value. The fixed percentage, *about 10% of the voting population,* was included as a strategy used in schools and in the wider community. The numeric values responses are arranged in descending order in decrements of an order of magnitude. The two populations were described as, for example, "1.5 million" and fully numerically, for example, "1,500,000." An additional comment encouraged students to focus on sample size by emphasising that a representative and random sample was taken. The statistically correct and commonly used sample size for a large population is, for both tasks, (c) 1500.

Students were asked to provide an explanation for their choice of sample size, with the question posed informally as "what best describes your thoughts". In both items students were presented with four explanations and the further option of volunteering a written response. The four explanations: "(i) I read it in a newspaper or heard it on TV" identifies the media as the students' source of information; "(ii) I mainly thought about the practicalities and cost of doing the survey" provides students with an opportunity to make a considered judgement amongst the alternatives presented; "(iii) eliminated a few then guessed" recognises a strategy used commonly; and "(iv) knew it from school" explores whether students had encountered the topic in school previously. In Task 2 students are presented with the same four explanations, and a fifth: "(i) the sample size has to be smaller because the population is smaller" identifies a commonly held belief of sample size.

Task 1: Opinion survey prior to an Australian national election

This question was much more topical prior to last year's federal election! In Australia's population of 20 million about 15 million (=15,000,000) are 18 years of age or older and can vote. A survey is to be conducted on "how people will vote at the next national election". How many voters do you think should be surveyed? [This is a question about sample size; assume that the sample is perfectly representative with men and women, and young and old voters etc. included in the sample in the same proportion as the entire voting population.]

(a) About 10% of the voting population (b) 15,000 (c) 1,500 (d) 150

Why did you choose that sample size? Circle what best describes your thoughts:

(i) I read it in a newspaper or heard on TV, (ii) I mainly thought about the practicalities and cost of doing the survey, (iii) eliminated a few then guessed, (iv) knew it from school, (v) other:

Task 2: Opinion survey prior to a Queensland state election

In a Queensland state election there are about 1.5 million (=1,500,000) voters. How many voters should be surveyed for a Queensland state election?

(a) About 10% of the voting population (b) 15,000 (c) 1,500 (d) 150

Why did you choose that sample size? Circle what best describes your thoughts:

(i) the sample size has to be smaller because the population is smaller (ii) I mainly thought about the practicalities and cost of doing the survey (iii) I read it in a newspaper or heard on TV (iv) eliminated a few then guessed (v) knew it from school (vi) other:

Results

Consistent with the two test items' structure the analysis is performed in two parts: *Part A: Students' naïve sampling strategies* are presented as Tables 1-5, and *Part B: Students' explanations for their naïve sampling strategy* are presented as Table 6.

Part A: Students' Naïve Sampling Strategies

Table 1 presents students' naïve beliefs of sample size for a large population. Greater than half of both the boys and the girls chose an incorrect and impracticable sample size of 10% of the population of 15 million. The frequency of students' responses broadly descended in decreasing order of sample size, with the second most favoured strategy a sample size of (b) 15,000. Seven (21%) girls gave the (preferred) response of (c) 1,500, but in this multiple-choice item the result differs little from chance. No student gave the response of (d) 150.

Table 1
Students' Response to Task 1: Opinion Survey for an Australian National Election

Student response	Boys *(n=21)*		Girls *(n=33)*	
	Number	%	Number	%
(a) About 10% of the population	13	62 %	18	55 %
(b) 15,000	6	29 %	6	18 %
(c) 1,500[1]	2	9 %	7	21 %
(d) 150	0	0 %	0	0 %
No response	0	0 %	2	6 %
Total	21	100 %	33	100 %

[1] Note: The accepted sample size when sampling from large populations

A traditional analysis would examine students' response to the companion Task 2, but a more productive alternative is to examine whether students adopted a consistent sampling strategy response to the two tasks. Within this study, a consistent strategy is considered to be either a *percentage strategy* where students choose *about 10% of the population*, or a *numeric strategy* where students choose explicitly a numeric value (but not necessarily the same numeric value), for both tasks; for example, (b) 15,000. The term *consistent strategy* is used in the sense of students' responses: students' thinking may be consistent, but the two test items may not reveal this thought process. *Inconsistent strategies* are a combination of percentage and numeric strategies. Table 2 shows that in both classes 67% of students applied a consistent strategy, and 28% of boys and 15% of the girls used an inconsistent strategy. A substantial proportion of the girls did not provide a complete response to both tasks.

Table 2
Students' Use of Consistent or Inconsistent Sampling Strategies

Student response	Boys *(n=21)*		Girls *(n=33)*	
	Number	%	Number	%
Consistent sampling strategy	14	67 %	22	67 %
Inconsistent sampling strategy	6	28 %	5	15 %
Incomplete response to both items	1	5 %	6	18 %
Total	21	100 %	33	100 %

Students who adopted a consistent approach provided a sub-group of the cohort that may be divided further into three groups: (a) students who adopted a consistent percentage strategy, (b) students who used a consistent numeric strategy and (c) students who used the same sample size for both tasks. Table 3 shows that the female students preferred (68%) the percentage strategy, and the male students were approximately evenly divided between the use of a consistent percentage (43 %) and a consistent numeric (50%) strategies. One student from each class chose the same, and the preferred, sample size of 1,500 for both populations.

Table 3

Type of Consistent Strategy Used by Students

Response	Boys *(n=14)*		Girls *(n=22)*	
	Number	%	Number	%
Consistent percentage	6	43 %	15	68 %
Consistent numeric with a smaller sample used for the smaller state population	7	50 %	6	27 %
Consistent numeric; same sample size used for both national and state elections[2]	1	7 %	1	5 %
Total	14	100 %	22	100 %

[2] Note: Two students gave the preferred response of a sample size of 1500 for both items.

Students who adopted an inconsistent strategy comprised the complementary, albeit small, sub-group of six male and five female students. The small number of students in the sub-group provides interesting supporting information. The students predominantly *first* chose a percentage strategy for Task 1, *then* a numeric strategy for Task 2, as shown in Table 4.

Table 4

Type of Inconsistent Strategy Used by Students

Response	Boys *(n=6)*		Girls *(n=5)*	
	Number	%	Number	%
Percentage for Task 1, numeric for Task 2	5	83 %	4	80 %
Numeric for Task 1, percentage for Task 2	1	17 %	1	20 %
Total	6	100 %	5	100 %

In Task 2 the explanations included the alternative "(a) the sample size is smaller because the population is smaller". Responses to this item are confounded because it is impossible to distinguish between students who disagreed (the preferred response) and students who did not give any response. A large proportion of students, over half of the males and a third of the females, indicates this belief is held widely, as shown in Table 5.

Table 5

Students' Responses to Item Task 2 (a) "the sample size has to be smaller because the population is smaller"

Response	Boys *(n=21)*		Girls *(n=33)*	
	Number	%	Number	%
Agree	11	52 %	11	33 %
Disagree[3] / No response	10	48 %	22	67 %
Total	21	100 %	33	100 %

[3] Note: Preferred response.

This belief cannot be called a misconception because it is strictly correct; it is, however, a partial understanding only; survey design conventionally chooses a sample size determined by the accuracy required. For example, increasing sample size may increase survey accuracy, but such an increase in accuracy may not provide any additional meaningful information. It is the trade-off between sample size and the accuracy of the survey, and the ability to interpret and *make sense* of sample size that underpins good survey design.

Part B: Students' Explanations for Their Naïve Sampling Strategy.

Although students were asked to provide separate explanations for their strategies for the two tasks, the low level of responses for Task 2 means that responses to Task 1 are provided only. Arguably students' responses to the first task, Task 1, may represent their intuitive responses, summarised in Table 6. Similar proportions of male and female students offered "practicalities and cost" as the principal consideration. Over half of female students, but a smaller proportion of male students, provided the candid explanation of "eliminated a few then guessed." Almost one quarter of male students volunteered the written separate explanation "use a large a sample as possible to improve a survey's accuracy", but one female only offered this explanation. One student only gave the response "knew it from school", indicating that students had not encountered this topic at school previously. One male and one female identified the media as their source of information, but of these only one gave the preferred response of (c) 1,500.

Table 6
Task 1, Students' Explanations for their Strategies: "What best describes your thoughts?"

Response	Boys *(n=21)*		Girls *(n=33)*	
	Number	%	Number	%
Knew it from newspaper or TV	1	5 %	1	3 %
Considered practicalities and cost	6	29 %	10	30 %
Eliminated a few and guessed	3	14 %	17	52 %
Knew it from school	1	5 %	0	0 %
Other: accuracy, take largest sample	5	24 %	1	3 %
Other: intuition	2	9 %	0	0 %
Other: unclassified	0	0 %	1	3 %
No response	3	14 %	3	9 %
Total	21	100 %	33	100 %

Discussion

The two test items provoked a broad range of student responses. The responses provided by the male and female students differ substantially on whether a consistent strategy was used (Table 3) and on the explanation for the strategy used (Table 6), but these differences may reflect the education experiences of the two groups, rather than be gender based. Several themes emerged in the study, and these will be discussed in turn.

Part A: Students' Naïve Sampling Strategies.

A sampling strategy of "10% of the population" is used extensively, with over half over the students preferring this strategy. The sampling strategy's widespread appeal is obvious: it is simple to both remember and use, but the strategy is impracticable with the populations purposefully chosen for the two test items.

A consistent sampling strategy is preferred with two-thirds (67%) of both male and female students applied a consistent sampling strategy, either percentage or numeric, to both tasks. If applied consistently a percentage strategy is arguably a less sophisticated strategy than a consistent numeric approach because the "10% rule" may be applied without sensible consideration of the sample size required – a rule applied by rote. In contrast, and if purposefully used, a constant numeric approach is a two-step process that first requires an appreciation of the magnitude of the sample size (do students of this age have an intuitive sense of 150,000 people?), and second, the practical physical and financial considerations of conducting a survey – a modest demonstration of sense-making. The two students who chose correctly the preferred sample size of 1,500 could have used facts reported in the media.

The small group of students who adopted an inconsistent strategy (six male and five female), strongly favoured a percentage strategy for the first task, and a numeric value for the second task. It is interesting to speculate on students' thinking: perhaps they began with a familiar strategy then modified the strategy believing that a small sample size was required for Task 2, without appreciating that continued application of the 10% rule will achieve the same purpose.

Part B: Students' Explanations for Their Naïve Sampling Strategy

Students' naïve explanation for their preferred strategy represents their established and intuitive beliefs, and this may represent a point of departure for classroom instruction and a guide for students' development of understanding. Students provided predominantly three explanations for their sampling strategies. First, the high proportion (52%) of female students who explained their strategy as "eliminated a few then guessed" is an candid acknowledgment that students had little or no background knowledge to justify their sampling strategy. Second, almost a third (29% of male and 30% of female students) purposefully considered the practicalities and cost of conducting the survey. This suggests students are applying sense making to an everyday situation outside the mathematics classroom – sense-making that requires cultivation in the classroom. Third, 24% of male students favoured a large sample size to improve accuracy, which provides the opportunity for an exploration of sample size and accuracy. Reconciling the practicalities and the cost, and the accuracy of the survey, are precisely the same issues considered by professional statisticians when designing a survey.

That a smaller population does not require a smaller sample size is a counter-intuitive notion – after all, the converse of more people surveyed, the closer the result must approach that of the population seems axiomatic. Indeed this belief is formally correct, but the value of additional information gained through increasing the sample size may be inconsequential. This sophisticated notion is not normally encountered until tertiary level mathematics, but the foundation intuitions are potentially accessible to high school students.

Conclusion and Implications for Teaching and Research

The concept of sample size is somewhat neglected in education research and in schools, and many students may complete formal schooling with notions of sample size possibly first acquired at upper primary school. More sophisticated notions of sample size are potentially accessible to high school students by relating sample size and accuracy. The two test items, by quantifying high school students' beliefs of sample size when sampling from large populations, contribute modestly to the body of education research knowledge and may provide a formal mechanism to prompt and support further research in schools on this topic.

The two test items – prototypes only – require refinement. The explanation should include an option of "largest sample size possible to improve accuracy". In the second task "sample size should be smaller because the population is smaller" does not allow students who disagree to be distinguished from those who did not respond, so this question should be presented separately along with the responses agree/disagree/not certain. The explanations for students' choice of sample size should be offered once only at the conclusion of the test items.

Teaching entirely *to the two test items* misses the opportunity to explore both the relationship between sample size and survey accuracy, and the more sophisticated concept that the accuracy of survey in a large population is determined by the numerical sample size, not the sample proportion. Sampling is measurement; any measurement has a certain accuracy, and accuracy should be introduced first as sense-making of the familiar physical properties of mass, length, and time. The "10% of the population" sampling strategy, in common with any model, has limitations. This sampling strategy is clearly impracticable in the large populations used in the two test items, and it cannot be used with *infinite* populations that are encountered at school, such a die and coin systems – in the latter instance the sample size is chosen without any sound mathematical basis, but on the time available or the endurance of students to roll a die or flip a coin. The limitations of the "10%" strategy are potentially accessible to students in terms of time and cost of conducting a survey, or as an extension of the law of large numbers activities, or by the use of electronic simulations. Naïve notions identified by the two tasks may provide a basis for instruction. The challenge for teachers may be to replace naïve notions with more sophisticated sample size strategies.

Acknowledgment.

This research was supported under an ARC Linkage grant LP0669106 by the Australian Bureau of Statistics and KCP, the manufacturer of Fathom. The researchers thank the students and staff at the schools, Bill Finzer of Key Curriculum Press, and the first author's supervisor for their support and participation.

References

Australian Bureau of Statistics. (2010). *CensusAtSchool.* Retrieved 8 January 2010 from http://www.abs.gov.au/websitedbs/cashome.nsf/Home/Home

Department of Education, Tasmania. (2008). *The Tasmanian curriculum. Mathematics–numeracy. K – 10 syllabus and support materials.* Retrieved 2 November 2009 from http://resources.education.tas.gov.au/item/edres/29da4473-d8fd-6edd-a1d5-7e6da3384886/1/syl-mnall.pdf

Fielding, A. (1996). Determining adequate sample sizes. *Teaching Statistics*. *18*(1), 6–9. Retrieved 1 January 2008 from http://www3.interscience.wiley.com/cgi-bin/fulltext/119953047/PDFSTART

Smith, M. H. (2004). A sample/population size activity: Is it the sample size of the sample as a fraction of the population that matters? *Journal of Statistics Education. 12*(2). Retrieved 29 January 2008 from httop://www.amstat.org/publications/jse/v12n2/smith.html

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