Developing Concepts of Sampling
Author(s): Jane M. Watson and Jonathan B. Moritz
Published by: National Council of Teachers of Mathematics
Stable URL: http://www.jstor.org/stable/749819

REFERENCES
Linked references are available on JSTOR for this article:

You may need to log in to JSTOR to access the linked references.
Developing Concepts of Sampling

Jane M. Watson and Jonathan B. Moritz, University of Tasmania, Australia

A key element in developing ideas associated with statistical inference involves developing concepts of sampling. The objective of this research was to understand the characteristics of students’ constructions of the concept of sample. Sixty-two students in Grades 3, 6, and 9 were interviewed using open-ended questions related to sampling; written responses to a questionnaire were also analyzed. Responses were characterized in relation to the content, structure, and objectives of statistical literacy. Six categories of construction were identified and described in relation to the sophistication of developing concepts of sampling. These categories illustrate helpful and unhelpful foundations for an appropriate understanding of representativeness and hence will help curriculum developers and teachers plan interventions.

Key Words: Children’s strategies; Cognitive development; Intermediate/middle grades; Qualitative methods; Statistics; Stochastics

How do children develop the intuitions that form the foundation for sampling in statistics? Many educators appear to assume students understand the part-whole relationship of sample to sampled and the purpose of seeing what the whole is like. Is there a need to make more explicit the transition from out-of-school connotations of sampling, such as in supermarkets or medical contexts in which variation is not usually an issue, to representative sampling required for inference, sampling that must depict the underlying population without bias? This question reflects one of the educational issues underlying the research reported in this article. Another issue is associated with the goal of students’ achieving, before they leave school, a level of statistical literacy that will allow them to contribute meaningfully to social decision making based on quantitative data. In the context of sampling, meeting this goal requires that students develop both an understanding of sampling methods and the motivation to be able to question claims based on biased sampling methods.

CURRICULUM BACKGROUND

The importance of sampling to statistical analyses is stated explicitly in a general statement on statistical inference in A National Statement on Mathematics for Australian Schools (Australian Education Council [AEC], 1991):

The dual notions of sampling and of making inferences about populations, based on samples, are fundamental to prediction and decision making in many aspects of life. Students will need a great many experiences to enable them to understand principles underlying sampling and statistical inference and the important distinctions between

This research was funded by the Australian Research Council, Grants A79231392 and A79800950. The authors thank anonymous referees for helpful suggestions.
a population and a sample, a parameter and an estimate. Although this subheading [Statistical Inference] first appears separately in band C [secondary school], the groundwork should be laid in the early years of schooling in the context of data handling and chance activities. (p. 164)

The foundations for statistical inference are written into the Australian upper primary school curriculum; experiences are suggested to enable students to “understand what samples are, select appropriate samples from specified groups and draw informal inferences from data collected” (AEC, 1991, p. 172). Possible activities for building these foundations include (a) “discuss and decide how to select a simple random sample (e.g. by drawing names out of a hat, or by giving each child a number)” and (b) “discuss when it is appropriate to use a sample rather than a census of all the population” (p. 172). For students moving through the middle school years (Grades 5–8) in the United States, similar expectations are expressed in the National Council of Teachers of Mathematics’ (NCTM) Curriculum and Evaluation Standards for School Mathematics (1989):

Sampling procedures are a critical issue in data collection. Which students should be surveyed to determine Mr. or Ms. Average? Must every student be questioned? If not, how can randomness in the sampling be assured and how many samples are needed to accumulate enough data to describe the average student? Random samples, bias in sampling procedures, and limited samples all are important considerations. (p. 106)

This development of the concept of a sample as some but not all of the population prepares students for the curriculum, in later school years, that addresses issues of sample size and selection methods for representative samples to avoid bias. In Australia, by about Grade 10, high school students should be able to “understand what samples are and recognize the importance of random samples and sample size, and draw inferences and construct and evaluate arguments based on sample data” (AEC, 1991, p. 179). Similarly in the United States, during Grades 9 to 12, “students must acquire intuitive notions of randomness, representativeness, and bias in sampling to enhance their ability to evaluate statistical claims” (NCTM, 1989, p. 169).

As well as specifying content for developing concepts of sampling, the mathematics curriculum documents explicitly state the need for students to be statistically literate participants in society. The NCTM, for example, claims that “knowledge of statistics is necessary if students are to become intelligent consumers who can make critical and informed decisions” (1989, p. 105). In New Zealand the Ministry of Education (1992) reflects this view in devoting one of the three strands of its statistics curriculum to “interpreting statistical reports,” which includes at Level 6 “evaluating statistics in the news media, and in technical and financial reports” (p. 193). Similarly in Australia students are expected to “understand the impact of statistics on daily life” (AEC, 1991, p. 178). These statements imply that the concepts taught in the statistics curriculum, including sampling, should be considered not only as basic concepts for data collection projects in mathematics classes but also as concepts relevant in social contexts, to be applied by those reading claims made in the media and elsewhere.
RESEARCH BACKGROUND

Although the curriculum documents set objectives for developing concepts of sampling, there has been little research into the early development of students’ sampling cognitions. Reviewers of statistics education research (Shaughnessy, 1992; Shaughnessy, Garfield, & Greer, 1996) have noted with concern students’ difficulty in accepting variability in populations, unwarranted confidence in small samples, and lack of understanding that the size of random samples is important. These issues are noted as emerging research themes with no completed research mentioned.

Most research related to sampling has taken place with college students and has evolved from the early work of Tversky and Kahneman. They began their investigations of statistical reasoning with a detailed account of the widespread belief in the “law of small numbers” (1971), whereby people “expect that the essential characteristics of the process will be represented, not only globally in the entire sequence, but also locally in each of its parts” (1974, p. 1125). This belief was identified as part of a commonly used representativeness heuristic, the tendency to assume that a sample, irrespective of its size, represents the population (Kahneman & Tversky, 1972). Only 20% of 95 Stanford undergraduates, when asked whether a large hospital averaging 45 births per day or a small hospital averaging 15 births per day would more often have more than 60% boys born, correctly chose the smaller sample as more often yielding unrepresentative results. Success rates were only slightly better for contexts of average word length in a printed document and heights of men in a survey. The researchers concluded, “The notion that sampling variance decreases in proportion to sample size is apparently not part of man’s repertoire of intuitions” (p. 444). These are complex questions that involve comparison of the tails of the sampling distribution of large and small samples. When college students were asked whether a large or small sample is more likely to represent the population, success rates improved markedly (Well, Pollatsek, & Boyce, 1990).

Little is known about school students’ appreciation for sample size in either complex or simple settings. Fischbein and Schnarch (1997), however, included the “hospitals” problem in a study of probabilistic intuitions of students in Grades 5, 7, 9, and 11 ($n = 20$ each). Only one student chose the small hospital as likely to have the more extreme result, and the misconception of equal likelihood for the two hospitals grew steadily from Grade 5 to Grade 11. A similar result was obtained in the context of tossing coins 3 times versus 300 times. Fischbein and Schnarch suggested that as the understanding of ratio improves over the grades, it becomes dominant at the expense of understanding of the effect of sample size.

Recently attention has been given to the understandings of primary and secondary students concerning variability in samples. Wagner and Gal (1991) investigated third- and sixth-grade students’ ideas on sampling in interviews about comparing two data sets under different conditions. They found that responses depended on students’ conceptions of variation within data sets, whether they assumed homogeneity or anticipated natural variation. Working with 12 senior high school
students, Rubin, Bruce, and Tenney (1991) also found students’ appreciation for sample variability an important consideration, creating a tension with the need for sample representativeness. Mokros and Russell (1995) documented 21 students’ increasing awareness of representativeness in data sets leading to an appreciation of the need for a measure of average to represent a set. They considered students’ construction of an underlying variable distribution when given only the mean but did not specifically address the issue of sample in this context. The issue of bias related to various sampling methods has been considered by Jacobs (1997) and by Schwartz, Goldman, Vye, Barron, and The Cognition and Technology Group at Vanderbilt (1998). In both studies upper elementary school students preferred methods that, in their view, were fair in terms of opportunity for selection (such as voluntary participation) but that were statistically biased, and they often did not recognize the statistical fairness of a random sample.

The expressed concern noted in the curriculum documents for students’ understanding of statistics in everyday life is beginning to be acknowledged in research focusing on statistical literacy. Researchers recognize the need to transfer school-based understanding into contexts like those provided by the news media and to help students develop a questioning attitude when such is warranted (Gal, 1997; Watson, 1997). Within this wider context of statistical literacy, media reports associated with sampling in social settings have provided situations in which to observe developing understanding (Watson, 1997, 1998).

CONCEPTUAL FRAMEWORK

Two related perspectives contribute to the conceptual framework used in this study: the mathematical content and the contextual complexity associated with the application of content understanding. Through the mathematical-content perspective, we address the statistical appropriateness of responses in terms of curriculum objectives. Using the contextual-application perspective, we evaluate responses in terms of a hierarchy for statistical literacy. Although these perspectives are described separately in this section, they are considered in relation to each other in the subsequent data analysis.

Mathematical Content

In seeking to establish the ways in which students construct an understanding of samples and sampling, one must make judgments against relevant content criteria. Although the significant ideas are noted in the curriculum documents cited earlier, authors of many introductory statistics books appropriate for secondary schools do not address them, preferring to concentrate on graphical data representation (e.g., Graham, 1987; Landwehr & Watkins, 1986) or instructions for survey developers (Jackson, 1989). The importance of sampling to David Moore, however, is shown in his liberal arts text (1991) that devotes chapter 1 to sampling and begins as follows: “Boswell quotes Samuel Johnson as saying, ‘You don’t have
to eat the whole ox to know that the meat is tough.’ That is the essential idea of sampling: to gain information about the whole by examining only a part” (p. 4). The rest of the chapter deals with issues of sampling related to selection, bias, and applications to social surveys.

In the context of a text focusing on surveys, Orr (1995) introduced the term population followed by a short section on sample.

Sometimes, for theoretical, practical, or efficiency reasons, it is desirable to study (collect data from) less than the entire population. In such cases a subset of the population called a sample is selected. Although data are then collected only from or about the sample, conclusions are drawn (generalized) to the larger population as well. . . . What is the essential nature of a sample? In a word, a sample should be “representative.” This means that, effectively, a sample should be a small-scale replica of the population from which it is selected, in all respects that might affect the study conclusions. (p. 72)

These characteristics encompass the aspects of the concept of a sample considered important in this study. Except for passing comments related to appropriate sample size, however, expanded discussion of this issue in most college texts is placed in the context of a more detailed consideration of sample variance (e.g., Moore & McCabe, 1993). At the high school level, Landwehr, Swift, and Watkins (1987) are atypical in discussing methods of sampling, sources of bias, and sample size. The most comprehensive basic discussion on sampling issues appropriate for middle schools found appears in the Used Numbers book for Grades 5–6 (Corwin & Friel, 1990), which includes specific teacher notes on “samples, populations and predictions” and “taking a sample.” The goal for the unit with these teacher notes is “to have students understand … the relationships between samples and the populations from which they are drawn” (p. 16). Descriptions of a sample and discussions of sample size, method of selection, and sources of bias provide the basis required for later years, and it is these components that are to be explored in this study.

*Application in Context*

Evaluating students’ success in satisfying the demands of statistical literacy in context, as well as considering mathematical content, is important. A model for evaluating students’ responses was provided by Watson (1997), who posited three tiers in a hierarchy, the highest level of which represents the objective for students when they leave school. The achievements associated with the tiers are as follows:

- **Tier 1.** A basic understanding of statistical terminology;
- **Tier 2.** An understanding of statistical language and concepts when they are embedded in the context of wider social discussion;
- **Tier 3.** A questioning attitude one can assume when applying more sophisticated concepts to contradict claims made without proper statistical foundation.
In Tier 1, skills are associated with the basic language of topics in the curriculum. Although, in general, topics like mean, median, odds, measures of spread, or graphing would be included at this level, the emphasis in this tier is on sampling and associated ideas of representativeness. For example, a Tier 1 question such as “What is a sample?” might evoke a response such as “A sample is a little part of something to show what it is like.” Tier 2 includes skills associated with recognizing and applying the basic language in wider contexts. Often the context is social, but it may be scientific; examples of such contexts abound in the media. Recognizing that a survey reported before an election represents only a sample of the opinions of voters and being able to appreciate the meaning of confidence limits put on an estimate of voter support are examples of Tier 2 understandings. At the Tier 3 level, sophisticated skills allow for the questioning of statistical claims when such questioning is appropriate. In the preelection-survey example, the sample size might not be reported or the sample size might be too small to allow for confidence to be placed in the estimate. Students operating in Tier 3 will be alert to such situations. This model has been useful in assessing students’ responses to media articles based on pie graphs that do not sum to 100% (Watson, 1997), to tasks on graphing variables in a suspicious cause-effect relationship (Watson, in press), and to tasks designed to illustrate media claims based on unrepresentative samples (in the newspaper articles described as part of the current study, Watson, 1997, 1998).

Having a model for planning and assessing experiences involving statistical literacy fits into the overall contexts of literacy and numeracy. With respect to literacy, there has been much debate on approaches to critical literacy in relation to the learning of basic literacy skills. In one model of “reading as social practice” (Luke & Freebody, 1997, p. 214), which parallels the model for statistical literacy presented here, four relevant practices are advocated. The first practice develops resources of the learner as a code breaker, understanding terminology as in Tier 1. The next two develop resources of the text participant and text user, understandings incorporated in Tier 2. Finally, the last practice develops resources of the learner as a text analyst and critic, comparable to the questioning attitude in Tier 3. With respect to numeracy, Joram, Resnick, and Gabriele (1995) addressed the issue of “numery as cultural practice” in their analysis of the occurrence of fractions, percentages, and averages in magazines for children, teenagers, and adults. In their coding scheme they recognized distinctions among factual statements, conditional statements, claims, and opinions using numbers. Many of the examples presented include aspects of statistical literacy and hence could be used to motivate students or to assess all levels of the hierarchy above.

METHOD

Participants

The participants for this study were 62 students in Grades 3, 6, and 9 (ages 8–9, 11–12, and 14–15 years, respectively). For each grade, students were selected from
three or four different school regions, including suburban and rural, within the Australian state of Tasmania; approximately equal numbers of females and males were included. The distribution of students by grade is shown in the first row of Table 1. All students had taken part in one or two large-scale written surveys of concepts in chance and data (Watson, 1994); students were selected for interview to represent a range of levels apparent in survey responses; also included were some students who gave interesting or unusual responses.

Table 1  
Number of Respondents, by Grade, Completing Each Survey Question and Interview Part

<table>
<thead>
<tr>
<th>Item</th>
<th>Grade 3</th>
<th>Grade 6</th>
<th>Grade 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survey Q1</td>
<td>21</td>
<td>21</td>
<td>20</td>
</tr>
<tr>
<td>Survey Q2</td>
<td>—</td>
<td>21</td>
<td>20</td>
</tr>
<tr>
<td>Survey Q3</td>
<td>—</td>
<td>—</td>
<td>18</td>
</tr>
<tr>
<td>Interview Part 1</td>
<td>21</td>
<td>21</td>
<td>20</td>
</tr>
<tr>
<td>Interview Part 2</td>
<td>18</td>
<td>21</td>
<td>20</td>
</tr>
<tr>
<td>Interview Part 3</td>
<td>3</td>
<td>19</td>
<td>19</td>
</tr>
</tbody>
</table>

Note. Students in Grades 3 and 6 were not administered some survey questions.

Items

The items from the written surveys included in this analysis are shown in Figure 1. Item Q1 was administered to all participants; Item 2 was administered to Grade 6 and Grade 9 students. Because of its sensitive subject matter, Item 3 was administered to Grade 9 students only.

Students were interviewed to explore in more detail their understanding of sampling; the protocol is presented in Figure 2. The interviews were designed following the Collis-Romberg problem-solving format (Collis & Romberg, 1992) so that each successive part required more complex cognitive functioning (Biggs & Collis, 1982). In Part 1a, a student was meant to reiterate views concerning the term sample and its basic meaning. Parts 1b and 2 were designed to introduce, within a concrete context, aspects of sample definition, purpose, size, and selection. Part 3, designed to investigate students' understanding of the effect of increased sample size in increasing the reliability with which the sample represents the population, was adapted from previous research (Kahneman & Tversky, 1972). The context was changed from births in hospitals to children in schools to conform to the context of earlier parts of the protocol concerning sampling Grade 5 children to determine weight. Whereas the original "hospitals" problem was modeled by a binomial distribution (p = .5) for the birth of a boy or a girl, the "schools" problem involved a hypergeometric distribution because each child can be selected for a sample only once from the finite school population, that is, without replacement. None of the students involved in this study had had any exposure to the theoretical bases of these two models, and it was not expected that they would base reasoning on such
Q1. If you were given a "sample," what would you have?

Q2. ABOUT six in 10 United States high school students say they could get a handgun if they wanted one, a third of them within an hour, a survey shows. The poll of 2508 junior and senior high school students in Chicago also found 15 per cent had actually carried a handgun within the past 30 days, with 4 per cent taking one to school. (The Mercury, 21 July 1993, p.17)

(a) Would you make any criticisms of the claims in this article?
(b) If you were a high school teacher, would this report make you refuse a job offer somewhere else in the United States, say Colorado or Arizona? Why or why not?

Q3. Decriminalise drug use: poll

SOME 96 percent of callers to youth radio station Triple J have said marijuana use should be decriminalised in Australia. The phone-in listener poll, which closed yesterday, showed 9924—out of the 10,000-plus callers—favoured decriminalisation, the station said.

Only 389 believed possession of the drug should remain a criminal offence. Many callers stressed they did not smoke marijuana but still believed in decriminalising its use, a Triple J statement said.

(The Mercury, 26 September 1992, p. 3)

(a) What was the size of the sample in this article?
(b) Is the sample reported here a reliable way of finding out public support for the decriminalisation of marijuana? Why or why not?

Figure 1. Sampling items from written surveys.

considerations. Because the conclusion obtained in either situation would suggest that the smaller sample size is more likely to be extreme, it was felt that responses would not be affected by the difference in models. In fact, no students discussed this matter of nonreplacement.

Procedure

Students in this study had participated in a statewide survey of 2187 students (Watson & Moritz, 1998); the written surveys had been administered to whole-class groups during mathematics class time. Item Q1 was from a 20-item chance and data survey; Items Q2 and Q3 were from a media survey (Watson, 1994). Students were selected for interviews on the basis of the variety and sometimes the unusual nature of their responses to survey items. Selections were not necessarily based specifically on responses related to sampling. Teachers confirmed that the students selected by the researchers were able and willing to be interviewed. One or the other of us interviewed individual students in a separate room for 45 minutes during class time; all interviews were videotaped. Students were told that they could stop the interview at any time, but none chose to end the interview early. The questions in Figure 2 constituted a protocol that was similar to nine others used during the interview,
1. (a) Have you heard of the word sample before? Where? What does it mean?
   (b) A news person on TV says, "In a research study on the weight of Tasmanian Grade 5 children, some researchers interviewed a sample of Grade 5 children in Tasmania." What does the word sample mean in this sentence?

2. (a) Why do you think the researchers used a sample of Grade 5 children, instead of studying all the Grade 5 children in Tasmania?
   (b) Do you think they used a sample of about 10 children? Why or why not?
   (c) How many children should they choose for their sample? Why?

3. The researchers went to 2 schools:
   - 1 school in the centre of the city, and 1 school in the country.
   - Each school had about half girls and half boys.
   - The researchers took a random sample from each school:
     - 50 children from the city school,
     - 20 children from the country school.
   - One of these samples was unusual: It had more than 80% boys.
   - Is it more likely to have come from
defined situations?
   - □ the large sample of 50 from the city school, or
   - □ the small sample of 20 from the country school, or
   - □ are both samples equally likely to have been the unusual sample?
   - Please explain your answer.

Figure 2. Three parts of the interview protocol for sampling.

which covered various aspects of the chance and data curriculum. The protocol was followed closely, although particular responses sometimes triggered other probing questions asked to assess more fully students' understandings. The questioning about sample was occasionally cut short when the interviewer felt that the student was becoming uncomfortably confused or frustrated. Table 1 shows the numbers of students at each grade level who answered each survey question and interview part. Most Grade 3 students were not asked Part 3 because earlier responses indicated that the students could not reasonably be expected to justify a response choice.

Analysis

Using the language analysis software NUD*IST (Qualitative Solutions and Research, 1995), students' (n = 2187) written responses to the survey items had previously been classified, on the basis of the Structure of Observed Learning Outcomes model of Biggs and Collis (1982), into levels (Watson & Moritz, 1998). For the purpose of this analysis of the selected 62 students, three levels of thinking are identified. At the unistructural level, students focus on single elements required to solve a task presented. In describing a sample, for example, students give a single characteristic, sometimes involving an example or one defining feature, such as "a little bit" or "something to test." At the multistructural level, students mention
several aspects associated with the set task, usually in sequence. Responses to “What is a sample?” include a reference substance or “whole” of which the sample is a part or a test, such as “a little of something, not the whole thing but a little piece of it” or “I would have a sample of dirt, something they have done tests on.” At the relational level, students refer to all elements required to solve a task in a coordinated fashion, describing a small part representing a whole. Examples include “a small portion of something larger to try” or “I would have a piece of something to show me what the whole thing looks like; for example, a carpet.” If conflicting ideas are expressed by the student, at the unistructural level the conflict is not recognized, at the multistructural level it is recognized but without resolution or closure, whereas at the relational level it is resolved. These three levels provided a structure to define increasingly sophisticated responses, not only to the survey items but also to the first two parts of the interview protocol.

For the interview, a checklist was developed on the basis of earlier pilot work (Watson, Collis, & Moritz, 1995) to cover commonly occurring responses to each part. The pilot study was based on responses of 30 students in Grades 3 to 9 (none of whom participated in the larger study) using the same protocol as in Figure 2. We used this check list, independently, to code responses when viewing all interview videotapes. Any discrepancies in our interpretations of what had been said were resolved during joint viewing of the tapes. In our preliminary analysis, we considered responses to each part of the protocol separately. Using this approach, we found inconsistencies in response levels for many students, for example, when students offered no response to a question they had responded to earlier. We felt that responses needed to be considered in relation to the whole understanding of the student as evidenced by responses to all parts of the interview and the written survey questions. We studied interview transcripts to explore subtleties of student reasoning not always apparent in checklists and to identify categories of common reasoning.

The criteria for placing students in categories were based on the statistical appropriateness of the content in the contexts provided by the interview protocol and media survey items. Although most students appeared to offer quite similar descriptions of a sample, the structural complexity of responses was a distinguishing feature. For choice of sample size, for Part 2 we asked students to respond to a sample size of 10. Students who considered 10 an appropriate size or suggested a sample of fewer than 15, often for idiosyncratic reasons, were classified as small samplers. Responses of 20 or more were usually associated with a recognized need for more information and appropriate selection method and were classified as large samplers. Because there were no responses suggesting a sample size between 15 and 20, we believe that in most cases this was a reasonable distinction. Methods of selection were associated with predetermined of the characteristic (e.g., “pick some fat and some thin”), with random choice, or with a primitive stratification procedure involving distribution from different schools (e.g., “pick some from each school in the state”). Students could demonstrate recognition of bias in four contexts: by identifying potential bias in the selection of children to study weight
(Interview Part 2), by choosing the small sample from a school (Interview Part 3), by questioning the sample from Chicago in the claim about gun use in the United States (Survey Q2), and by making one of several appropriate criticisms of the phone-in poll (Survey Q3) (e.g., “only youth listen,” “only interested people call,” or “callers may be lying”).

The methodology employed to categorize responses closely resembles the technique of clustering described by Miles and Huberman (1994). This technique was carried out on the basis of the conceptual framework and the procedures described in the previous paragraphs applied to the survey and interview responses. The

<table>
<thead>
<tr>
<th>Small Samplers Without Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>• may provide examples of samples, such as food products</td>
</tr>
<tr>
<td>• may describe a sample as a small bit or, more rarely, as a try or test</td>
</tr>
<tr>
<td>• agree to a sample size of fewer than 15</td>
</tr>
<tr>
<td>• suggest no method of selection or an idiosyncratic method</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Small Samplers With Primitive Random Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>• provide examples of samples, such as food products</td>
</tr>
<tr>
<td>• describe a sample as either a small bit or a try or test</td>
</tr>
<tr>
<td>• agree to a sample size of fewer than 15</td>
</tr>
<tr>
<td>• suggest selection by random without description or with a simple instruction to choose “any,” perhaps from different schools</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Small Samplers With Preselection of Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>• provide examples of samples, such as food products</td>
</tr>
<tr>
<td>• describe a sample as both a small bit and a try or test</td>
</tr>
<tr>
<td>• agree to a sample size of fewer than 15</td>
</tr>
<tr>
<td>• suggest selection of people by weight, either a spread of fat and skinny or people of normal weight</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Large Samplers With Random or Distributed Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>• provide examples of samples, such as food products</td>
</tr>
<tr>
<td>• describe a sample as both a small bit and a try or test</td>
</tr>
<tr>
<td>• may refer to term average</td>
</tr>
<tr>
<td>• suggest a sample size of at least 20 or a percentage of the population</td>
</tr>
<tr>
<td>• suggest selection based on a random process or distribution by geography</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Large Samplers Sensitive to Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>• provide examples of samples, sometimes involving surveying</td>
</tr>
<tr>
<td>• describe a sample as both a small bit and a try or test</td>
</tr>
<tr>
<td>• may refer to terms average or representative</td>
</tr>
<tr>
<td>• suggest a sample size of at least 20 or a percentage of the population</td>
</tr>
<tr>
<td>• suggest selection based on a random process or distribution by geography</td>
</tr>
<tr>
<td>• express concern for selection of samples to avoid bias</td>
</tr>
<tr>
<td>• identify biased samples in newspaper articles reporting on results of surveys</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Equivocal Samplers</th>
</tr>
</thead>
<tbody>
<tr>
<td>• provide examples and descriptions of samples</td>
</tr>
<tr>
<td>• may indicate indifference about sample size, sometimes based on irrelevant aspects</td>
</tr>
<tr>
<td>• may base decisions on small size with appropriate selection methods or with partial sensitivity to bias, or base decisions on large sample size with inappropriate selection methods</td>
</tr>
</tbody>
</table>

**Figure 3.** Characteristics of six categories of developing concepts of sampling.
production of a spreadsheet with descriptive categories associated with the definitions of sample (with 7 subgroups), sample size (6 subgroups), methods of selection (7 subgroups), and bias (6 subgroups) allowed for student columns to be moved to sort for visual clustering of designated cells in the matrix as well as for pattern identification within clusters (Miles & Huberman, 1994). Six categories of developing concepts of sampling were identified, as summarized in Figure 3. The first five categories reflected increasing use of appropriate mathematical content and complexity of structure in responses, ranging from limited or no experience with samples to a demonstrated ability to question claims that were based on results from biased sampling procedures. Students in the Equivocal Sampler category offered responses that could not be assigned to one of the five previous categories; these students are discussed separately.

RESULTS

The distribution of responses for the three grade levels over the six categories is shown in Table 2. In the first five categories, all Grade 3 students were small samplers whereas all Grade 9 students were large samplers, with Grade 6 students offering the most diverse range of responses. An expected trend for higher level performance with increasing age is seen. In the following presentation of results we describe each category and provide examples of student responses. All extracts include enough of the interviewer’s wording of the question to indicate the meaning found in Figure 2. Use of “…” denotes a pause in the student’s response, and “[... ]” denotes dialogue that has been edited. For the sake of brevity and for highlighting certain features of responses, only selected responses are presented. When a student’s responses from both the survey and the interview are provided, note how often the student offered very similar responses to each, although interviews were conducted a number of weeks after the surveys had been administered.

<table>
<thead>
<tr>
<th>Category</th>
<th>Grade 3</th>
<th>Grade 6</th>
<th>Grade 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small Samplers Without Selection</td>
<td>12</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Small Samplers With Primitive Random Selection</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Small Samplers With Preselection of Results</td>
<td>5</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>Large Samplers With Random or Distributed Selection</td>
<td>0</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>Large Samplers Sensitive to Bias</td>
<td>0</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>Equivocal Samplers</td>
<td>0</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>
Small Samplers Without Selection

Twelve Grade 3 students offered responses indicating that they were developing a basic concept of sample without any clear consideration for size or selection. In describing a sample, most of these students identified an example from their experiences and described a sample as “a small bit.” When asked how many children should be chosen in a sample for a study of the weight of Grade 5 children, all chose a size of fewer than 15. Only 5 of the 12 students offered any ideas for selection, and these were idiosyncratic ideas based in experience of the context of children in the classroom, without applying the notion of sampling to this context.

Students who could not recall anything about the word sample sometimes went on to describe aspects of the context compatible with sampling. One Grade 3 student appeared to have gleaned from the context the cue of a research study in which children were interviewed; on the basis of this cue, she suggested that a sample is related to “studying something on them.” This idea is possibly a primitive element students learn to associate with sample before they develop the idea of a sample as a test to show or represent something.

Most students offered their own examples of samples that were based on experiences, often involving food products or something related to science or medicine. The student who responded in the following extract appeared to know something about samples from experience with free samples of products; she knew the defining characteristic that a sample is small. The student considered acceptable the sample size of 10 and, on the basis of the context, made an idiosyncratic suggestion for selection.

S1: (Grade 3) [Survey Q1: If you were given a sample, what would you have?] Something free. [Interview Part 1a: Meaning?] A little packet or something. Like something free, a little packet or something. [Part 1b: Meaning in context of weight of Grade 3 children?] Like, just had some. [Part 2b: How many? 10?] Probably about 10. [Part 2c: How choose?] Teacher might just choose people who’ve been working well or something.

Sometimes students’ reliance on their own experiences resulted in inappropriate responses for the context of the task. In the following responses, a student described an appropriate example of a sample product from her environment but then, for the context provided, continued to use her own personal experience.

S2: (Grade 3) [Survey Q1: If you were given a sample, what would you have?] You would be trying something. [Interview Part 1a: Meaning?] I’ve tried some food. At [the supermarket] you get to try some food [...] they cook something or get it from the shop, and they put it in a little container for you so that you can try it. [Part 1b: Meaning in context of weight of Grade 5 children?] They might take a blood test on them.

As students develop a concept of sample on the basis of their experiences, they appreciate that a sample is a bit of something and that it provides a test. This finding is illustrated by the following student who used these notions in very simple expressions to respond to the task.

S3: (Grade 3) [Survey Q1: If you were given a sample, what would you have?] An example of a real medicine. [Interview Part 1a: Meaning?] In medicine, a little sample of a big

Overall these 12 Grade 3 students who appeared to be developing a basic understanding of the term sample, limited to the first tier of the statistical literacy hierarchy, gave remarkably similar responses. Beyond recognizing a sample as a few or some of the Grade 5 children, these students showed no evidence of applying a sampling concept involving testing from a small group in the context to inform appropriate decisions related to size or selection.

Small Samplers With Primitive Random Selection

Five students who offered only partial descriptions in defining a sample and selected small sample sizes went on to suggest an early developing concept of random or stratified selection methods. Three of these defined a sample in terms of a test and two defined it as “a small bit”; no students affirmed both a test and a bit. All offered an example, again often involving a food product or a medical or a science setting.

One Grade 3 student described a sample as “a little bit,” then expressed an idea about randomness, but this idea was not linked to a need for a sample to be representative of a population.

S4: (Grade 3) [Interview Part 1a: Meaning?] A blood sample is taking a little bit of blood. [Part 1b: Meaning in context of weight of Grade 5 children?] Some of the people. [Part 2a: Why not all?] Because they didn’t have enough time. [Part 2b: How many? 10?] Umm, could have used 10. [Why?] It’s an even number and I like that number. [Part 2c: How choose?] Go by random. [Why?] Because they’re not really worried about what people they pick.

The following Grade 3 student described the idea of a test from science and transferred this understanding to the context of sampling Grade 5 children. The student expressed the idea that sampling is the process of testing them, but this idea appeared to be unconnected with the selection of “a couple from each school” for the sample.

S5: (Grade 3) [Interview Part 1a: Meaning?] I suppose it’s like when scientists are studying a certain thing, and they’re taking a sample of the thing so they can study it. [Part 1b: Meaning in context of weight of Grade 5 children?] They probably took a couple of Grade 5 children and weighed them and then they took a sample of children to use for their research. [Part 2b: How many? 10?] Yes, probably about 10 children ... [Part 2c: How choose?] Probably by taking sort of like a test and then, sort of like, choosing a couple of people who made sense with their work. And then just, they choose a couple from each school ... probably got umm ... [end of response].

The five students in this category commonly exhibited simple, unistructural responses to individual questions throughout their interviews. Although these students had some isolated ideas about choosing from schools or classes or had heard the word random, they could not explain their meanings or relate them adequately to the context. Their acceptance of a sample size of 10 indicates that
they had developed a very limited appreciation of the requirements of a sample. Although these students’ ideas about sampling went beyond those of students in the first category, they would still be considered within the first tier of statistical literacy, developing terminology.

In developing sampling concepts involving the importance of size and selection for representativeness, one must acknowledge the possibility of variation in the population. Indeed, if there were no variation, then size and method of selection would be irrelevant inasmuch as any sample of any size would show the same thing. Students in the first two categories gave no indication that they appreciated variability; all their examples were of substances expected to have high uniformity.

**Small Samplers With Preselection of Results**

Acknowledging weight variation in the population, 11 students suggested selecting a sample of children by their weights, either to ensure that both skinny and fat children were in the sample or to exclude this variability and select only children who appeared to be about the normal weight. These students failed to realize that selection by weights would predetermine the results and, hence, compromise the purpose of a sample: to see what the whole is like.

Students in this category also chose small sample sizes and offered examples of samples of apparently homogeneous substances. In describing the meaning of a sample, however, they gave more thorough responses than those in previous categories in that all 11 students included both “a small bit” and “a test or try” somewhere in their descriptions. The following Grade 3 student’s responses illustrate these features. The idea of testing, initially of dirt, was later applied in the context of weights as a purpose for comparison, perhaps involving the concept of average.

S6: (Grade 3) [Survey Q1: If you were given a sample, what would you have?] Some dirt to test. [Interview Part 1a: Meaning?] If you sample something, you take a ... sample ...; like if you were looking at dirt, you’d find out just what was in it, if there were toxics or not. That’d be sampling dirt. [Part 1b: Meaning in context of weight of Grade 5 children?] Take out a few of them, I’d say. Not many. [Part 2b: How many? 10?] They could use any number, I’d say. Ten would be the [...] number I’d use [...]. [Part 2c: How choose?] I would choose them in all shapes and sizes, some skinny, some fat. Then I’d compare them to another group and see what was the most average.

Another Grade 3 student gave an elaborate idea of sampling by weight. Although she appeared to have forgotten it was weight and not height that was being measured, she clearly understood that the sample was being used to show what the whole was like. To demonstrate this understanding, she first defined the whole as 20 or 30 students in the class and then described how individual people of different heights [weights] could be used as representatives of others.

S7: (Grade 3) [Interview Part 2c: How choose?] If they saw a really small person, bring them up and then measure some of the other members of the class with them that they also thought were small. And if there are, the other people are, around their height, they could choose them. And if there was a really tall person, they could measure some of the others with them, and they’re around that height. Then you could take those two.
This student appears to be expressing a primitive idea of representativeness: to ensure that all types of weight are present in the sample.

Some students in this category also advocated selecting the sample by weight, but they would choose only children of the normal weight for their samples. They acknowledged variability of weights and the purpose of testing to find the normal weight, but they wanted to preselect the normal weight (presumably by appearance) instead of selecting for a spread. The following Grade 6 responses provide an example.

S8: (Grade 6) [Interview Part 1a: Meaning?] In a shop [...] I like try it before you buy it to see if you like it [...]. [Part 1b: Meaning in context of weight of Grade 5 children?] They took some Grade 5 children and just checked to see their weight [...] like try and see how much they weighed. [All?] Well, just about a few I suppose, just to work out, work out probably the normal weight a Grade 5 should be, like if they were overweight or underweight or something. [Part 2a: Why not all?] They probably took some people who looked not overweight, just they looked a good weight [...] a right weight for their age [...]. They probably only took a few because a few of them looked the right weight.

Students in this category generally demonstrated that they had developed a basic concept of a sample as a small part to show what the whole is like. They also appeared to be moving into the second tier of the statistical literacy hierarchy in taking on a notable aspect of the context, variation. Students who selected normal weights assumed that the study’s aim was to find the average weight of children in the state, whereas those who selected different weights may have assumed similar aims but perhaps also believed that the researcher needed to find the distribution or variation of weights. These students’ attempts to grapple with the variability in the weights of Grade 5 children were more sophisticated and structurally complex than those of the previous groups, though the students were unsuccessful in terms of resolving potential bias and recognizing the importance of sample size.

**Large Samplers With Random or Distributed Selection**

Fifteen students demonstrated an understanding of the need for appropriately selected larger samples, but they did not mention potential bias in sampling. The two large-sample categories were different from the small-sample categories in respects other than just size. All *large* samplers offered methods of selection, either by random methods or by selection from schools in different locations. In comparison to the Small Samplers With Primitive Random Selection, these Large Samplers With Random or Distributed Selection offered more developed expressions of selection methods, and their descriptions of a sample were comprehensive, including both the “small bit” and “test” ideas that were not present together for any of the Small Samplers With Primitive Random Selection. Many students in this group of *large* samplers also included a notion of average in connection with the goal of the research study. Students choosing a larger sample were thus identified as displaying relational thinking in successfully combining several aspects in their descriptions of concepts of sampling.
The following responses from a Grade 9 student demonstrated both the need for using a larger sized sample to determine the average weight for the population and a method that involved selection from each school.

S9: (Grade 9) [Interview Part 1a: Meaning?] People give you, like if there’s any kind of cheese or something like that, people give you a little bit to try it. [Part 2b: How many Grade 5 children? 10?] They would need more than that to sort of find out what sort of average weight is for Grade 5 students. [How many?] 100, about. [Why?] Usually they have a survey of 100 people in most things. [Part 2c: How choose?] Pick a few schools, then just have 5 out of each school, 10 out of each Grade 5 so it’s just sort of evened out... not all from one school, but a few different schools with only a few from each Grade 5 class.

It is interesting to note how the expression of “some out of each school” compares with the idea of “some of each weight type” observed for responses in the previous category. Acknowledgment of variation was implied in both, but a large distributed sample was suggested in this response to “even out” the variation instead of predetermining this by selection according to weight. A Grade 6 student emphasized the size of the population and elaborated on how to implement random selection.

S10: (Grade 6) [Interview Part 1b: Meaning in context of weight of Grade 5 children?] [...] About 10,000 children in Grade 5 in Tasmania and they took about say 50 of them and they found out, because 50 is a small portion of 10,000, they just found out the weights from there. [Part 2c: How choose?] Say take a kid from each school. Take some—just pick a kid from random order. Look up on the computer; don’t even know what the person looks like or anything. Pick that person.

The students in this group recognized the need for a larger sample size and appropriate methods of selection in the context of researching the weights of Grade 5 children in the state. They thus appeared to be developing concepts of size and selection of samples applied in context, within the second tier of the statistical literacy hierarchy. These students either created no conflict or resolved any conflict to achieve a coherent explanation in the context set in the protocol. Significantly, however, only two of these students recognized the likelihood for an extreme result to come from the small sample in Part 3 of the protocol, whereas three others showed limited recognition of the bias in the sampling methods used in one of the two newspaper articles in the survey. None of these students had hence reached the third tier of the statistical literacy hierarchy by consistently recognizing bias in more than one context.

Large Samplers Sensitive to Bias

Students in this category included in their responses all the features of the previous category as well as the additional feature of recognition that representativeness and avoiding bias are important issues in sampling. Twelve students offered responses demonstrating a comprehensive understanding of examples and defining characteristics of a sample: the need for large and random or distributed selection methods together with recognition of the potential for bias in sampling.
in at least two settings. These students often linked these ideas to the purpose of sampling as representing the whole or to the average as a representative measure.

Although recognizing potential bias, students in this category showed varying degrees of confidence in expressing their ideas about samples. The following responses are from a Grade 6 student who was developing concepts of representativeness and bias in relation to sampling. She failed to recognize bias in response to the first question about the Chicago sample (Survey Q2a) but acknowledged the variation in location when prompted (Q2b). She also noted potential bias in the sample of weights when asked about sample size (Part 2b of the protocol).

S11: (Grade 6) [Survey Q2b: Would you refuse a job somewhere else?] The report would not make me refuse a job in Colorado or Arizona, but I would not take a job in a more dangerous place like Chicago or L.A. [Interview Part 2b: How many Grade 5 children? 10?] Probably some more, because if they only used 10, they could all be ones that weighed about the same, and there could be some that weighed less and weighed more in other places. [How many?] Probably about 100 or something. [Part 2c: How choose?] Just choose anybody; just close your eyes and pick them or something.

One Grade 9 student, in her survey responses, demonstrated a concern for representation by referring to random selection (Q3b) and expressed a concern that every student be represented (Q2a and Q2b), albeit with a dubious notion of size required for sampling. In the interview, this student expressed an understanding of the possible bias with the phrase “say they got all fat kids” and noted that a larger random sample would be likely to result in a distribution representative of the population.

Students choosing large sample sizes often used the word random and explained it carefully.

S12: (Grade 9) [Interview Part 2c: How choose?] Just randomly; you don’t want to look at them; you just want to get a computer screen in the office [...] just randomly choose them all, just to make it fair. If you’re going to choose all the fat children, then it’s going to put the average right up, isn’t it? And if you want to be fair, [...] you’d randomly choose them like that.

The most complete responses reflected the third tier of statistical literacy in terms of a consistent appreciation of characteristics of good samples across contexts. The notion of a cross-sectional method of selection and the potential for bias permeate the following responses of a Grade 9 student.

S13: (Grade 9) [Survey Q2a: Any criticisms of the claims about guns in United States schools?] No, not really, because it is probably true, but the poll should have taken a larger cross section. [Survey Q3b: Is the sample reliable concerning public support for the decriminalisation of marijuana?] No, because a lot of people would not be bothered ringing in. Plus only the really active supporters would, due to the law being against them. [Interview Part 1a: Meaning?] Sample is like try something or a small part of something like in surveys and stuff; a sample is like a person from here, there, and everywhere, just a sample or like a cross section of a community or whatever, or like at the supermarket—they give away samples like a piece to try. [...] [Part 1b: Meaning in context of weight of Grade 5 children?] They interviewed like a cross section, all different students from here, there, and everywhere all round Tasmania. They didn’t like use every single student, [...] just randomly picking them. [Part 2b:
How many? 10?] I think they would use a lot more than 10 because if they’d used 10, it could be a lot under or a lot above because you could have extremely high results. [...] [How many?] It depends how many Grade 5 children there are in the state. I don’t know, perhaps 10% or 20% of Grade 5 students in the state. [Part 3: Larger or smaller sample for unusual result?] I think it [you] would be more likely to get it [the sample of 80% or more boys] from the country school because there are a lot less children, so if you had perhaps a few more, you would bring the percentage up a lot quicker than with this sample [city school] [...].

Students in this category expressed an awareness of the problem of bias in sampling in more than one context and thus were in the third tier of the statistical literacy hierarchy. Of these 12 students, all mentioned sample bias in the weight context; 7 acknowledged bias in one of the media survey questions, whereas 5 mentioned it in both media articles; of these 5 students, 3 also appreciated the relation of size to bias in Part 3 of the interview. Hence, although some still appeared to struggle to construct and apply their relational concept of sample in all settings, others consistently applied their sense of bias in various contexts to question claims.

Equivocal Samplers

Seven students offered responses that were in some sense equivocal in terms of the choice of sample size. It was therefore impossible to assign the responses to one of the five previous categories. For several students an apparent tentative choice associated with one of the small- or large-sampler categories was mitigated by a method of selection that was more typical of the other group.

Three students made contradictory suggestions that indicated an indifference or confusion about sample size, which made assigning them to either small- or large-sampler categories impossible. In one response, the dominant idea was that sample size did not matter, although the student suggested 20 or 30 “so that’s just like a sample.” Another student was more concerned about divisibility of the sample size into the total and hence did not appreciate the underlying issue of sample size.

S14: (Grade 6) [Interview Part 2a: Why not all children in Grade 5?] It’s sort of like an average. They’re just taking a few of them and then checking them. Doing the whole lot would be very difficult. So that’s why they’re only using a smaller sample; then you know they kind of multiply those amounts by 10 or something. [...] They get, kind of get, an average weight from just a small sample of them. [Part 2b: How many? 10?] I reckon they’d use probably 10 because rounded, say there are 15,000 or something, they round things, and they’d want to have a number that could divide equally into that; [...] it could have been 10 or 100; I reckon they would use 10 or a multiple of 10.

The third student did not resolve conflicting ideas about representativeness and variability of samples. He responded not only that any number was appropriate but also that the whole population should be tested to be accurate.

S15: (Grade 9) [Interview Part 1b: Meaning in context of weight of Grade 5 children?] A few Grade 5s. [Part 2a: Why not all?] They should have really [asked all] because a sample of the population is not true of Tasmania. It doesn’t mean you get a good sample. [Part 2b: How many? 10?] ... Doesn’t necessarily have to be 10; could be any number.
[Should use all or a few?] Should use a lot ..., most of the Grade 5 children in Tasmania. [Any advantage to a few rather than all?] Well you wouldn’t know the average, would you, if they only sampled a few? So probably better off with the whole.

One Grade 6 student tentatively accepted a small sample size but then displayed more sophisticated reasoning on other questions.

S16: (Grade 6) [Interview Part 2b: How many Grade 5 children? 10?] Yes, that would be all right, I suppose. [Part 2c: How choose?] Just choose anybody, really. [How?] Just pick them out. [Part 3: Larger or smaller sample for unusual result?] I think 20, because 80%, you wouldn’t have as many to take 80% from ... [Interviewer asks for more explanation] Yes, because say 50 boys and 50 girls, they just take anybody; there’s more chance of being more boys than girls if there’s only 20.

In contrast to this Grade 6 student, another, focusing on variation, showed support for a larger sample size of about 40 because “with 10 they could be fat, skinny” but suggested a selection method of “all different weights” characteristic of pre-selection of results. Finally two Grade 9 students suggested random methods of selection and thought that large sample size was important but that 10 or 14 was sufficiently large. These were the only two instances in which an explicit connection was made between such small numbers and the need for a “large” sample.

In relation to the hierarchy of categories suggested in Figure 3, Equivocal Samplers appeared to be in transition between the small samplers and the large samplers. All responses displayed multistructural complexity: Conflicting issues were not consistently resolved in a statistically appropriate fashion. Overall, however, the Equivocal Samplers displayed an understanding typical of students in the second tier of statistical literacy.

DISCUSSION

Categories of Developing Concepts of Sampling

Six categories of developing concepts of sampling were identified in students’ responses to a variety of survey and interview questions, as shown in Figure 3. The first five categories form a hierarchy of increasing sophistication concerning sample size, selection, and resulting representativeness. The hierarchy of categories thus may be hypothesized as a model of student development of concepts of sampling. Students initially build a concept of sample from experiences with sample products or in medical- and science-related contexts, perhaps associating the term random with sampling. As students begin to acknowledge variation in the population, they recognize the importance of sample selection, at first attempting to ensure representation by predetermined selection but subsequently by realizing that adequate sample size coupled with random or stratified selection is a valid method to obtain samples representing the whole population. As valid methods of sampling are consolidated, sample data are interpreted with appreciation of how sample size and selection contribute to biased or representative samples.

The six categories of response can be seen to fit within the three tiers of the hierarchy for statistical literacy, with the Equivocal Samplers in the middle of Tier 2.
Tier 1  Small Samplers Without Selection
           Small Samplers With Primitive Random Selection
Tier 2  Small Samplers With Preselection of Results
           Equivocal Samplers
           Large Samplers With Random or Distributed Selection
Tier 3  Large Samplers Sensitive to Bias

Tier 1 students are developing a concept of sample without application. They have a basic idea of the language, often expressed in ideas related to sample, and sometimes a context-free use of the idea of randomness. Tier 2 students have richer concepts of sample that can be applied in straightforward contexts, although some have not yet recognized appropriate selection techniques and some struggle with sample size. Tier 2 encompasses the largest range of students: some retaining small-sampling ideas but within a context in which the importance of sampling methodology is recognized, some apparently in transition with conflicting ideas unresolved, and some with well-developed sampling and selection methods but unable to go to the critical stage. Tier 3 students have a well-developed sense of sample that includes appropriate selection methods and the recognition of bias in many or all situations in which it can occur. Except for several students in the Equivocal Samplers group, students generally did not criticize biased samples without also having an appropriate understanding of sample size and selection. The results support the hypothesis that there is a hierarchical progression in the development of concepts of sampling.

This framework for describing developing concepts of sampling may be useful to teachers and curriculum planners who are interested in providing experiences to help students structure more complex responses and at the same time achieve higher levels of statistical literacy in social contexts. Two key concepts emerge as significant for developing concepts of sampling, each deserving separate discussion. One is the role of the appreciation of population variation in the transition from Tier 1 to Tier 2. The other is the development of sensitivity to bias in the transition from Tier 2 to Tier 3; related to this development is the question of why some students who advocate large random samples (Large Samplers With Random or Distributed Selection) do not question claims based on biased samples.

Developing Concepts of Population Variation and Representation

Concepts of variation and representation are central to both descriptive and inferential statistics. Statistics can be broadly defined as the mathematical study of variation. Graphical and tabular representations of data are important for conveying variation in data. Using averages that measure central tendency, one attempts to find a single representative measure of a data set, and measures of spread, such as variance and standard deviation, are fundamental to inferential statistics. Variation and representation are not surprisingly also key concepts in considering sampling techniques for data collection and making inferences about populations on the basis of sample data.
Evidence of the importance of the concept of population variation to students’ developing concepts of sampling comes most clearly from responses of students in the category of Small Samplers With Preselection of Results. Of the three categories of small samplers, this category appears to be the most developed inasmuch as students in this category provide the most complete descriptions of the basic concept of sample and are the clearest in expressing a method of selection. Methods involve either selecting both fat and thin children to represent the variation students know to be present in the population or selecting children of normal weight, avoiding fat and thin children. These seem to be students’ first attempts to handle population variation. These findings are similar to those from other studies: Rubin et al. (1991) found tensions between representativeness and variability, and Jacobs (1997) and Schwartz et al. (1998) observed students’ mistrust of random sampling and their preference, in order to ensure fairness, to intentionally sample each strata of the population or invite volunteers.

Students in the category of Small Samplers With Preselection of Results showed less sophisticated reasoning than those in large-sample categories. Apparently dealing with population variation prompts the need for larger sample sizes and also for random or stratified selection methods. Instead of focusing on the variation seen in fat and thin children as the Small Samplers With Preselection of Results did, Large Samplers With Random or Distributed Selection often referred to the population, describing the large number of children spread out across Tasmania. This attention to the population is related to the idea, expressed by Mokros and Russell (1995), of using average as a representative measure of a data set:

Until a data set can be thought of as a unit, not simply as a series of values, it cannot be described and summarized as something that is more than the sum of its parts. An average is a measure of the center of the data, a value that represents aspects of the data set as a whole. An average makes no sense until data sets make sense as real entities. (p. 35)

Similarly, it would appear that statistical samples make no sense until populations make sense as real entities. Variability in populations thus would appear to deserve explicit attention in statistics curricula. Research on the understanding of variability is also required (Shaughnessy, 1997).

**Developing Concepts of Sampling Bias**

Identifying sample bias is an important facet of statistical literacy, as recommended for higher grade levels in United States and Australian curricula (AEC, 1991; NCTM, 1989). A significant result of the current study is that a student’s demonstration of a preference for large random samples is not sufficient to ensure that the student will identify sample bias, as evidenced by the 15 Large Samplers With Random or Distributed Selection who did not acknowledge bias. Although the 9 Grade 6 students in this category attempted one survey question fewer (Q3) and so had fewer opportunities to demonstrate sensitivity for bias, they did fail to acknowledge bias in other questions. In addition, although all Large Samplers
Sensitive to Bias made some comment about potential bias in the interview context, 4 of the 12 made no mention of this for the Chicago item (Q2), even when prompted (Q2b). It appears that students’ success in identifying bias may be heavily dependent on contextual clues given.

Results from the larger study of survey responses, on the basis of which these students were selected (Watson & Moritz, 1998), support this view. For the question on access to guns in Chicago (Q2a), no more than 2% of Grade 6 students \((n = 506)\) and 5% of Grade 9 students \((n = 625)\) acknowledged the bias that the sample was only from Chicago. Even given the subsequent hint (Q2b), only a further 3% to 5% of students at each level went on to notice this bias. Results were less disappointing for the item concerning the phone-in supporting marijuana decriminalization (Q3), for which 30% of Grade 9 students \((n = 625)\) could give at least one reason the marijuana survey was biased. It may be that Australian students were more likely to accept the conclusion concerning access to guns in the United States because of common perceptions of the foreign country, whereas the local survey reached a conclusion less compatible with existing topic knowledge, leading to students’ questioning of its claims.

Only 8 of the 41 students who were asked about large and small samples from schools (Part 3 of the interview protocol) chose response (b), the small sample, and of these only 6 gave an adequate reason for the choice. One of the other 2 students could not explain, and the other displayed no real appreciation for the task at hand. This latter Grade 6 student said, “Because country schools, they don’t have many people and they probably took too many boys.” This type of response was also characteristic of many who chose the larger city school, saying, for example, “The big one because there’s more children to pick and choose from ...; there’s more children to get that 80% boys from.” The majority of students (61%) who answered the question said that both schools had the same chance of having 80% boys in their samples either because the process was random or because of the 50-50 split of boys and girls in each school population. Hence 81% of those who appreciated that a larger sample was needed to study the weight of Grade 5 children in Part 2 did not recognize that the smaller sample is more likely to give an extreme or biased result in the more complex setting of Part 3. Even 75% of those who specifically mentioned a bias, like “a small sample might all be skinny,” did not achieve success on Part 3. These percentages are similar to those reported earlier from samples of college students (Kahneman & Tversky, 1972). Compared to the results of Fischbein and Schnarch (1997) with school-aged students, a higher percentage in the current study could justify the correct choice of the small sample, and there was no difference between the Grade 6 (12/19) and Grade 9 (11/19) students in choosing the equal-likelihood option. Because this study did not employ a similarly structured question concerning means rather than extreme percentages, we cannot comment on the conjecture of Well et al. (1990) that students would perform better in relation to the more easily conceived statistic. Responses to Part 2 about the weight of Grade 5 children, however, indicate that students may appreciate that a larger sample will be more accurate, even if they cannot identify
the consequences of this for sample variability in Part 3. The complexity of the latter question may require more sophisticated cognitive functioning.

**Teaching Implications for Developing Concepts of Sampling**

As with other terms in mathematics, the term *sample* has a relatively straightforward meaning when met in out-of-school contexts (cf. Jacobs, 1997) and a more technical meaning that applies when it is desired to represent a population in an unbiased fashion in order to draw conclusions about that population. The transition from the former meaning to the latter involves reconstructing the notion that suffices in the supermarket. Because of the apparently homogeneous nature of examples such as food or blood, information about taste or presence of disease can likely be determined from a small sample just as easily as from a large amount. Experiences with small samples of food and other products do not help to generate appreciation for the variation present in many wholes or populations and for the consequent need for samples to be large enough and representative enough to show realistically what the whole is like. Explicit discussion about collecting sample data for which the measure, such as height or weight, obviously varies in the population may be important for developing concepts of sampling. A useful strategy could be to state directly for students in the middle school years the difference between a slab of cheese sampled for taste in the supermarket and a population of Grade 5 children in the state sampled to study their weights. The expectation in terms of variation and results are quite different in the two contexts. Further, the question of sample size is often not introduced in the context of beginning sampling (e.g., Graham, 1987; Moore, 1991), and although the appropriate methods of selection may implicitly require appropriate sample sizes for implementation, many students apparently would profit from explicit discussion of this point.

Most Grade 9 students in this study had some appreciation of the complex issues associated with statistical sampling. They could appreciate the need for larger samples and appropriate methods of selection but often did not possess the language to provide convincing arguments based on representativeness and avoiding bias. There is some risk that this situation will not improve at the next level of education if those teaching the link between variability and sample size in the formal expression of \( \sigma_x = \sigma / \sqrt{n} \) implicitly assume that students understand the importance of random sampling. Such instruction may encourage students to focus merely on sample size reported in statistical claims without identifying bias in the sampling procedure. Apparently at all levels the need to have representative samples requires reinforcement. In magazines popular with teenagers, examples that use voluntary response surveys to make claims about their readership or wider populations seem to be a good starting point to motivate discussion.

The suggestions for practice arising from this study fit well within the context of numeracy as a cultural practice described by Joram et al. (1995). In their conclusion they lamented the apparent lack of instruction from textbooks on the interpretation of rational numbers in the context of the discourse often found in maga-
azines, particularly when “the goal is explicitly to comprehend and interpret information in a passage rather than to solve a specific problem” (p. 359). They went on to suggest the use of everyday texts to promote numeracy in the classroom. Derry, Levin, Osana, and Jones (1998) made a similar point at the conclusion of their study of Grade 8 students: The development of “statistical reasoning does not necessitate computation, but always involves interpreting and reasoning about real-world problems with conceptual structures representing ideas such as probability, correlation, and experimental control” (p. 190). The outcomes of the current study support these points, particularly in relation to the understanding of sampling.

Building on the results of this study, we suggest that teachers and curriculum planners need to be aware of the following points in relation to sampling. Many students need help in the transition from understanding a sample as a test extract of a relatively homogeneous substance to appreciating the importance of variation in the samples and populations that will be the subjects of statistical investigations. Here the common usage of the word sample is a mixed blessing: It tells us part but not all of the story required to be statistically literate. The colloquial term offers a nonrepresentative sample of the statistical term! Further, as well as teaching appropriate methods for selecting samples, teachers must help students develop appreciation for situations in which bias can occur. Using examples from students’ personal experiences and the media should help motivate the questioning attitudes required of future citizens.

REFERENCES


Qualitative Solutions and Research. (1995). Non-Numerical Unstructured Data • Indexing Searching and Theorizing (NUD•IST) (Version 3.0.5) [Computer software]. Melbourne, Australia: LaTrobe University.


Developing Concepts of Sampling


Authors

Jane M. Watson, Reader in Mathematics Education, Faculty of Education, University of Tasmania, GPO Box 252-66, Hobart, Tasmania 7001, Australia; Jane.Watson@utas.edu.au

Jonathan B. Moritz, Research Fellow in Mathematics Education, Faculty of Education, University of Tasmania, GPO Box 252-66, Hobart, Tasmania 7001, Australia; Jonathan.Moritz@utas.edu.au