How Do Students Think About Statistical Sampling Before Instruction?

Victoria R. Jacobs

How would your students respond to the Raffle Scenario in Figure 1? What information about your students' knowledge would help you plan instruction for statistical issues related to the Raffle Scenario? This article highlights students' thinking and instructional implications from two studies that examined upper-elementary students' written and oral explanations as they responded to survey situations. Specifically, students were asked to reason about statistical issues related to how the participants were selected—the *sampling method*—for the following:

- *Random* sampling methods that gave each member of the population the same chance of being selected (for example, survey 1 in the Raffle Scenario) or each member within subgroups of the population having the same chance of being selected (for example, survey 2).
- *Restricted* sampling methods that asked particular groups of people who might be more likely to select a certain response and, consequently, bias the results in a particular direction, for example, surveys 3 and 4 in the Raffle Scenario.
- *Self-selected* sampling methods that had the participants select themselves and, consequently, might bias the results because individuals who choose to participate in a survey often have opinions different from the opinions of people who choose not to participate, for example, surveys 5 and 6 in the Raffle Scenario.

None of the students had formally studied sampling, but my goal was not to instruct the students on sampling issues. Rather, I wanted to find out what they already knew from their experiences outside of school---their informal knowledge.

Interviewer: What is a sample?
Melanie: A piece of something whole--it's like a peek.
Georgia: A piece of food or carpet [that] gives you an idea of what the real thing is.

Figure 1. Raffle Scenario (adapted from Schwartz et al. [1994]) for Examining Sampling Issues with Surveys

The Raffle Scenario

A Grade 5 class wants to raise money for a field trip to an amusement park. Students are considering several options, including selling raffle tickets for a videogame system. As part of their efforts to determine how to raise the most money, students conducted different surveys to estimate how many students in the whole school would buy a raffle ticket to win this prize. For each survey, 60 students were asked their opinion. The school consisted of 600 students in Grades 1--6 with 100 students in each grade. The surveys and their results follow:

1. Shannon got the names of all 600 kids in the school, put them in a hat and pulled out 60 of them. (Thirty-five percent said that they would buy tickets.)
2. Kyle put the names of all the first-grade boys in one hat and all the first-grade girls in another hat. He pulled out [the names of] five boys and five girls from each hat. He did the same thing for each grade until he had [the names of] five boys and five girls from each grade. (Thirty percent said that they would buy tickets.)
3. Raffi asked 60 of his friends. (Seventy-five percent said that they would buy tickets.)
4. Jake asked 60 kids at an after-school meeting of the games club. (Ninety percent said they would buy tickets.)
5. Claire set up a booth outside the lunchroom. Anyone who wanted to could stop by and fill out her survey. She stopped collecting surveys when she got 60 completed. (Ninety-five percent said that they would buy tickets.)
6. Abby sent out a questionnaire to every kid in the school and then used the first 60 that were returned to her. (Eighty-five percent said that they would buy tickets.)

What do you think about each of these survey methods? Good? Bad? Why?

What do you think is the best estimate of the percentage of kids in the whole school who will buy a raffle ticket?

Despite a lack of formal instruction, these students included some of the elements of statistical samples in their definitions: (a) the sample is part of the whole and (b) the smaller part gives an idea of the whole. This partial understanding could be a good starting point for instruction; teachers can build instruction on the basis of students' informal knowledge. This approach has proved successful in many areas of mathematics (Hiebert and Carpenter 1992). For instance, teachers have used students' experiences with fair sharing in out-of-school contexts, such as sharing four cookies with a sibling, as the basis for instruction on division with children as young as Kindergartners (Carpenter et al. 1993). Similarly, middle school teachers can benefit from understanding how upper-elementary students like Melanie and Georgia think about sampling issues before they receive formal instruction.

This article identifies students' informal knowledge of sampling by describing the major categories of their responses in two studies that asked them to evaluate individual surveys and then to draw conclusions from multiple surveys with conflicting results. In the first study, I interviewed 17 students to begin categorizing their conceptions of sampling. On the basis of a written assessment about sampling in surveys. I selected these students from 31 fourth graders and 32 fifth-graders in three multiage, fourth- and fifth-grade classrooms to include a range of understanding. In the second study, I developed a series of written activities from the categories of sampling conceptions identified in the first study. See Figures 2 and 3 for examples of these written activities. I asked 110 students to complete these written activities to confirm the categories identified in the first study and to determine the prevalence of the categories in a larger sample. These students came from eight fifth-grade classrooms in three elementary schools that were chosen to reflect the minority population—31.6 percent—and income distribution—29.2 percent of students receiving free and reduced-cost lunch—of this medium-sized city in Wisconsin.

Figure 2. Written Activity for Assessing Students' Abilities to Draw Conclusions from Multiple Surveys

	Will Buy Raffle Tickets	Will Not Buy Raffle Tickets
Shannon pulled 60 names out of a hat.	35%	65%
Claire set up a booth to collect 60 surveys.	95%	5%
Jake asked 60 kids in the games club.	90%	10%

What percentage of students in the whole school will buy a raffle ticket?

Here are some ideas that other students had. Circle any of the ideas that you agree with. Put a star next to the idea that you agree with most. If you do not agree with any of the ideas, circle the last choice and explain what you think.

A. I thought that Shannon's survey was the only one that was done well, so I ignored the other two surveys and used Shannon's results. She found that 35 percent said that they would buy a raffle ticket.

B. I thought that Claire's survey was the only one that was done well, so I ignored that other two surveys and used Claire's results. She found that 95 percent said that they would buy a raffle ticket.

C. I thought that Jake's survey was the only one that was done well, so I ignored the other two surveys and used Jake's results. He found that 90 percent said that they would buy a raffle ticket.

D. I just knew that most kids like video games and would buy a raffle ticket, so I picked a high percent. ____ percent of kids will buy a raffle ticket.

E. I just knew that most kids would not buy a raffle ticket, so I picked a low percent. ____ percent of kids will buy a raffle ticket.

F. I took the average of the three surveys. The average of the kids who said that they would buy a raffle ticket is 73 percent.

G. I don't know because they got different results.

H. I think ____ percent of kids in the whole school are willing to buy a raffle ticket because

_____.

Throughout this article, I illustrate categories with students' responses to the Raffle Scenario from interviews in the first study. I also indicate the prevalence of each response category by reporting percentages of the 110 students in the second study who used each response category in their evaluations of nine surveys. Percentages are not reported for the first study because they would not be meaningful. Each interview was individualized to investigate fully each student's thinking; therefore, different students had different opportunities to give particular responses.

Evaluating Individual Surveys

Each student evaluated the quality of individual surveys in scenarios like the Raffle Scenario, adapted from Schwartz and colleagues (1994). Other scenarios included statewide reporting on recycling programs, choosing classroom pets and identifying favorite lunchroom items. Some students showed sound reasoning by basing their evaluations on the potential of the sampling method for producing a biased sample—a group of people who would be likely to produce results that were not reflective of the population. Other students showed more problematic reasoning by basing their evaluations on other issues. On average across nine surveys, 50 percent of the students used sound reasoning; 47 percent of them used problematic reasoning; and 3 percent used reasoning that was missing, unique or unclassifiable. All students used multiple types of rationales when evaluating different surveys instead of sticking with a favorite rationale.

Sound Reasoning

Some students evaluated sampling methods by focusing on just what we would want them to consider: the quality of the sample and the potential for bias. This sound reasoning based on the potential for bias led students to make both accurate and inaccurate evaluations of surveys.

Accurate Evaluations

On average, 34 percent of the students used sound reasoning that led to accurate evaluations of survey quality. It is important to note that this sophisticated way of evaluating sampling methods was not restricted to a small group of "smart" students. Ninety-four percent of the students used this approach at least once when evaluating the nine surveys. With this approach, students positively evaluated random sampling on the basis of their tendency to produce unbiased samples, and they negatively evaluated restricted and self-selected sampling methods on the basis of their tendency to produce biased samples. For example, students negatively evaluated restricted sampling methods if they recognized that these methods were problematic because they were likely to produce samples in which everyone has the same opinion. In the Raffle Scenario, one student negatively evaluated the restricted sampling method of selecting only friends because—

> friends a lot of times are friends because they have the same opinions ... so a lot of his friends are going to like one thing or the other. And it seems to me they mostly like getting raffle tickets.

Similarly, students negatively evaluated self-selected sampling methods when they recognized that these methods were problematic because the individuals who choose to complete surveys are likely to have opinions different from those people who do not choose to participate. In contrast, students positively evaluated random sampling methods if they were able to recognize the potential for producing an unbiased sample. For instance, one student positively evaluated the stratified random sampling method of selecting five boys and five girls from each grade because—

> that way he has a mixture of boys and girls and who are different ages ... because sometimes girls and boys can have different opinions on things and also one age might really like something, but an older age might think that was a terrible idea.

Inaccurate Evaluations

On average, 16 percent of the students used basically sound reasoning, measured against the potential for bias, but this reasoning was applied inappropriately and consequently led to inaccurate evaluations of survey quality. For example, some students positively evaluated self-selected sampling methods. They incorrectly assumed that these methods would produce a good mixture of respondents because no sample restrictions were specified, such as selecting only girls or friends. The following student positively evaluated the self-selected sampling method of sending a questionnaire to every student in the school and then using the first 60 returned:

> I think Abby's was a good idea because ... she got a variety of people ... because you weren't just giving them to a couple people and then giving it back to you. You were asking every kid, and whoever wanted to return it, could.

Striving for a mixture is sound reasoning, but a mixture does not usually result from a self-selected sampling method.

Similarly, some students negatively evaluated random sampling methods when the mixture was not clearly specified, for example, putting names in a hat. They did not like the uncertainty as to who would be selected. One student suggested that with simple random sampling,

> you could get like all your friends, or all girls, or all boys, like all in the first grade or something and everybody else has different opinions.

Avoiding restrictions, such as all girls, is sound reasoning, but the likelihood of random sampling's producing a sample with all girls is very low. However these students seemed to focus on the *possibility* of extreme outcomes without realizing that the *probability* of their occurrence was low.

Problematic Reasoning

In contrast to the foregoing responses, some students in their evaluations showed more problematic reasoning by focusing on issues other than the potential for bias. These students based their evaluations on fairness issues, practical issues, results or all three.

Fairness Rationales

On average, 23 percent of the students inappropriately evaluated sampling methods on the basis of whether they were fair. However these students were not thinking of fair in the statistical sense of whether the sample would be fair, that is, whether everyone has an equal chance of being selected so that the sample is not biased. Rather, they were concerned about how people *feel* when they are selected or not selected to participate in the survey. Students using fairness rationales assumed that everyone wanted to participate in the surveys. To be fair, in an equitable sense, they believed that everyone should have the chance to participate.

Sometimes fairness rationales, and their problematic reasoning, led to accurate evaluations. For example, some students appropriately evaluated a restricted sampling method negatively, but their reasoning was based on the idea that the people who were left out would feel bad, not that the responses would be restricted and potentially biased by the people they selected. The following student negatively evaluated the restricted sampling method of selecting only friends because—

> that still wouldn't be fair. Because some people don't know him ... and they would say, "Hey, but this person told me that you picked them and not me. How come?"

In other situations, fairness rationales and their problematic reasoning led to inaccurate evaluations, especially in self-selected sampling methods in which everyone initially has a chance to participate. For example, one student positively evaluated the self-selected sampling method of setting up a booth for volunteers to complete the survey because—

> the people will choose if they want to. ... Like if they wanted to do the survey, they will, but if they would not want to, they don't have to—so they're not pressuring anybody.

For these students, the fact that everyone had a chance to participate—the students' idea of fairness—was more important than the fact that, using the self-selected method, people with particular opinions were more likely to participate than others.

Practical Rationales

On average, 12 percent of the students inappropriately evaluated sampling methods on the basis of whether actually conducting the survey would be practical. For example, was the sampling method efficient, easy to implement, confusing or even possible? This reasoning is not sufficient for evaluating the quality of sampling methods. In addition, students were not always accurate in their evaluations of which sampling methods were, in fact, practical. They would even sometimes suggest asking *everyone* instead of taking a sample, because they drastically underestimated the difficulties of asking everyone in large surveys, such as surveys of entire states.

Results-Based Rationales

On average, 12 percent of the students inappropriately evaluated sampling methods on the basis of whether the results were decisive, the results of the survey corresponded with their expectations or both.

First, some students based their evaluations on the decisiveness of the results. They believed that a survey with a completely decisive result—such as 100 percent of the students will buy raffle tickets—was more useful than a survey with indecisive results—such as 50 percent will buy tickets. These students concluded that a sampling method producing a 100 percent result was of a higher quality than a sampling method producing a 50-50 split because, as one student suggested, "50-50's not going to decide for you."

Second, some students based their evaluations on whether the results of the survey corresponded with their expectations of what would actually occur in the real world. When the results corresponded with their

expectations, these students evaluated the sampling method positively. Conversely, when the results did not correspond, they concluded that the sampling method was inappropriate. This finding is consistent with adult research that has found that we are more critical of ideas that are not consistent with our own (Lord, Ross and Lepper 1979). It is important to note that some students were able to separate their own evaluations of sampling methods from their own opinions *and* to articulate that separation. For example, one student commented on the stratified random sampling method of selecting five boys and five girls from each grade:

> I think that the way they picked the same number of boys and girls in each grade was a good way.... I don't know about those results, though.... They don't seem to really match what most of the kids I know would think.

Drawing Conclusions from Multiple Surveys

After evaluating the quality of individual surveys in scenarios like the Raffle Scenario, each student was asked to draw conclusions from these multiple surveys with conflicting results. See Figure 2 for an example of this type of activity. On average, 64 percent of the students drew their conclusions on the basis of survey results, whereas 31 percent did not. Five percent of the students' responses were missing, unique or unclassifiable.

Conclusions Based on Survey Results

On average, 20 percent of the students used survey results in a way that we would advocate. They initially evaluated the quality of the surveys and then used information only from the survey or surveys that they thought were done well (options A–C in Figure 2). For example, if students believed that Claire's and Jake's surveys were done poorly, they ignored those surveys and drew their conclusions from Shannon's survey (option A). Some students—on average, 44 percent—aggregated *all* the available information regardless of the quality of the surveys. In Figure 2, these students looked at all three surveys and computed a total, or average, percent of students who would buy a ticket (option F). This approach was particularly disappointing when students showed that they were capable of accurately evaluating individual sampling methods but then did not use this information when drawing conclusions from multiple surveys. Rather they aggregated all the information—often immediately after they had identified potential problems with some of the individual sampling methods.

Conclusions Not Based on Survey Results

On average, 31 percent of the students did not use the survey results to draw conclusions. Instead, some of them based their conclusions on other information, such as personal experiences or opinions (options D and E in Figure 2). Others essentially refused to draw conclusions from either their own experiences or survey results (option G in Figure 2), in particular, if the answer was not clear cut—for example, if one survey's results suggested that more students would buy a raffle ticket, whereas another survey's results suggested that more students would not buy a raffle ticket. As one student stated, "I can't think of what I would do if there were two different answers." Sometimes these students suggested conducting additional surveys to add support to one conclusion or another. At other times, they delegated the decision making to other individuals, such as authority figures like principals.

Assessing Students' Thinking

Teachers can most accurately assess their students' knowledge through oral questioning, such as interviewing, but this approach may be excessively time-consuming. Alternatively, teachers may want to use written activities similar to those used for the second study. Figure 2 is an example of a written activity to assess students' abilities to draw conclusions from multiple surveys. Figure 3 offers an example of a written activity to assess students' abilities to evaluate individual sampling methods. I developed both activities to correspond with the categories of students' thinking described in this article.

In Figure 3, option A reflects a problematic evaluation based on practical issues; option B, a well-reasoned evaluation based on the potential for bias; option C, a problematic evaluation based on results; and option D, a problematic evaluation based on fairness issues. Option E provides the opportunity for students to express an opinion in their own words. Given the time-consuming nature of interviews and the tendency to respond minimally to open-ended questions, this type of format may be a practical compromise to help teachers gain an initial picture of their students' understanding or identify which students' thinking they might need to explore in more depth. Whatever methods teachers choose to assess their students' thinking, they must probe for students' reasoning behind their answers.

Building Instruction on Students' Thinking

I found that even without formal instruction, upper-elementary students have substantial informal knowledge about sampling. Although this knowledge is not always complete, it can be a starting point for instruction. For example, instruction could begin with the basic premise of the fairness rationale: everyone should have an equal opportunity to participate. Instruction would then lead students to the realization that the importance of equal opportunity is to minimize the potential for bias in the resulting sample rather than to minimize the negative feelings of the nonparticipants.

Similarly, some students positively evaluated self-selected sampling methods because they inappropriately based their decisions on the idea that without stated restrictions, sampling methods are of high quality because they result in a mixture of people. Instruction should help these students retain this basically sound reasoning but recognize situations that require special consideration. For example, with self-selected sampling methods, everyone does not begin with the same probability of participating; therefore, even without stated restrictions, the resulting samples often do not include a mixture of people.

In addition to identifying students' specific conceptions, these studies suggest two considerations for designing instructional activities. First, teachers should give students opportunities to make decisions from the surveys. Students need practice in reasoning through decision making from surveys. It was not enough for students to evaluate individual surveys effectively, because many ignored their evaluations when drawing conclusions and making decisions. They often chose to aggregate the information from all surveys regardless of survey quality---even when they had already identified problems in some sampling methods.

Second, teachers should use surveys based in multiple situations. Students need experiences with situations both in and out of school. School situations are effective because they are familiar and interesting to students. However, out-of-school situations present more of a reason to sample because they have larger populations, which underscores the difficulty of asking everyone. It is important to note that a few students consistently evaluated sampling methods poorly because the students preferred to ask *everyone* rather than take a sample. However, students were less likely to insist on asking everyone if the survey referred to a situation that took place outside a school context. Experience with multiple situations should also help students base their evaluations on survey results rather than rely solely on personal experiences.

Figure 3. Written Activity for Assessing Students' Abilities to Evaluate Individual Sampling Methods

Raffi asked 60 of his friends, and he found that 75 percent said that they would buy raffle tickets and 25 percent said that they would not buy raffle tickets.

What do you think of Raffi's survey?

(Circle one) good bad I'm not sure

Here are some ideas that other kids had. Circle any of the ideas that you agree with. Put a star next to the idea that you agree with most. If you don't agree with any of the ideas, circle the last choice and explain what you think.

A. I made my decision because it was easy to do. He just had to ask people he already knew.
B. I made my decision because his friends probably agree with him. So the survey doesn't tell you how the people who are not friends with Raffi think.
C. I made my decision because most of the kids said they would buy raffle tickets.
D. I made my decision because it's not nice to the people who are not his friends. They want to answer the survey too, but they aren't allowed.
E. I made my decision because _____

_____

_____.

Conclusions

These studies suggest that students may be entering middle school with surprisingly rich informal knowledge about sampling that should be useful as a starting point for instruction. Therefore, although statistics instruction has traditionally been delayed until postsecondary education, my research convinced me that this delay is unnecessary. These results support the NCTM's (1989) recommendations that Grades 5–8 are appropriate times to introduce students to inferential statistics, of which sampling is a central component. This article identifies students' substantial, although incomplete, informal understanding of sampling in the transitional, upper-elementary grades, thereby underscoring the potential for instruction in middle school.

References

Carpenter, T., et al. "Models of Problem Solving: A Study of Kindergarten Children's Problem-Solving Processes." *Journal for Research in Mathematics Education* 24 (November 1993): 428–41.

Hiebert, J., and T. P. Carpenter. "Learning and Teaching with Understanding." *Handbook of Research on Mathematics Teaching and Learning*, edited by D. A. Grouws, 65–97. New York: Macmillan, 1992.

Lord, C. G., L. Ross and M. R. Lepper. "Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence." *Journal of Personality and Social Psychology* 37 (1979): 2,098–109.

National Council of Teachers of Mathematics (NCTM). *Curriculum and Evaluation Standards for School Mathematics*. Reston, Va.: NCTM, 1989.

Schwartz, D., et al. "Adolescent Understandings of Sampling in the Context of a Survey." Paper presented at the annual meeting of the American Educational Research Association, New Orleans, La., April 1994.

Reprinted with permission from Mathematics Teaching in the Middle School, *Volume 5, Number 4 (December 1999), pages 240–46, 263, an NCTM publication. Minor changes have been made to spelling and punctuation to fit ATA style.*

Multiplication Against Time

How is it possible to carry out the following multiplication in less than a minute?

81 624 324 048 566 472 808 896 × 12.5 =