Fourth Graders Engaged in Activities involving Sampling: A Case Study

Ko, Eun-Sung
Graduate School of Seoul National University, Department of Mathematics Education
1 Gwanak-ro, Gwanak-gu
Seoul 151-742, Korea
E-mail: kes-7402@hanmail.net

Park, Min-Sun
Graduate School of Seoul National University, Department of Mathematics Education
1 Gwanak-ro, Gwanak-gu
Seoul 151-742, Korea
E-mail:dpdlxl03@snu.ac.kr

INTRODUCTION

Common statistical information is not from populations but comes from samples instead. One should interpret the statistical information provided in the newspapers or on TV while minding the fact that the information is from samples. For this reason, it is needed to have a picture of surveying and to understand samples and sampling, crucial ideas related to surveys. Due to the importance of samples and sampling for statistical thinking, curriculum documents (e.g., Australian Education Council, 1991; Department for Education, 1999; Ministry of Education, 1992; National Council of Teachers of Mathematics, 2000) include recommendation to teach lower graders these topics. Studies (Fischbein & Schnarch, 1997; Shaughnessy, 1992) demonstrated that a variety of statistical ideas or concepts including samples and sampling are non-intuitive, even to adults. Researchers (e.g., Garfield, 2002) argue that students should form good intuition and a valid understanding of various statistical ideas or concepts through concrete activities and physical simulations including statistical contexts. This study examines fourth grade students engaged in three concrete activities involving sampling from finite populations. The analysis focuses on the challenges students encounter during their first experience with surveying or sampling and what contributions the activities make to students' thinking regarding surveys, samples and sampling.

BACKGROUND

Statistical inference is based on the basic condept that samples can provide valid information about the populations. However, the information from samples is not complete but approximate (Batanero, Godino, Vallecillo, Green, and Homes, 1994, p.527). Although we lose some of the information because we inference the information of populations using samples, we should try to obtain information that is as accurate as possible. Here, it is important to understand and to control the somewhat antithetical notions of sample representativeness and sampling variability (Rubin, Bruce, & Tenney, 1991, p.314). Sample representativeness reflects the idea that a sample is likely to have characteristics similar to those of its population when employing a proper sampling method (Batanero et al., 1994). Random sampling is part of the effort to produce samples representing their population by restraining bias. According to Saldanha and Thompson (2002), an understanding of sample representativeness comes by considering a sample as not a subset of the population but a quasi-proportional, small-scale version of the population. Watson and Moritz

(2000) also mentioned the importance of the understanding of relationship between random sampling and sample representativeness.

Studies have suggested that school students have difficulty with the ideas of sample and sampling. Watson and Moritz (2000) found that students put too much faith in a small sample, that is, the students believed that small samples represent their populations. Metz (1999) discovered that students argued the need to test all members of the population due to the variability in the population. Students in Jacobs (1999)' study preferred unbiased samples but they did not understand proper sampling methods. For example, they considered voluntary participation as a good sampling method due to fairness. Watson (2004, 2006) reported that students have difficulty in recognizing the importance of sample sizes in random sampling.

METHOD

The purposes of this study are twofold. One is to investigate what challenges students encounter during their first experience with surveys or sampling. The other is to examine what contributions the activities make to students' thinking as regards surveys, samples and sampling. To do this, the researchers develop three concrete activities involving sampling from finite populations: The first has students find methods to predict how many white and black marbles are in a box containing 300 marbles (Activity 1). The second requires students to design methods to predict how many times the Korean letter 'ga′ would appear in a Korean story book (Activity 2). The final activity included a survey of popular foods for school meals (Activity 3).

The participants in this study are four fourth graders (Choi, Jin, Kim, and Lee; all pseudonyms) who engage in a survey or sampling for the first time in their lives. They have learned only displays and summary of the given data using tables, bar graphs, and graphic charts. Activities were conducted as follows by one of the researchers as a teacher: A1 → A2 → A3 → A1 → A2, as the students were faced with many challenges during A1 and A2 on their first attempt. The teacher intended to give them opportunities to overcome the difficulties themselves through the experience from A3. Details are described subsequently. The students' thinking and behaviors were analyzed using video recordings, class observations, and the students' written work.

RESULTS AND DISCUSSION

Students' challenges

The teacher introduced Activity 1 to the students. She showed the students a box containing 300 marbles and had them find methods to predict how many white and black marbles were in the box. First, they were given some time to find the methods.

01 Teacher: How we can predict that? Anyone have any idea?
02 Lee: We can draw out one by one and count.
03 Teacher: That can be one way. Is there any other way?
04 Choi: Do you know that [how many white and black marbles are in the box] exactly?
05 Teacher: To predict means it's okay to say the approximate number. You do not necessarily have to Say the exact number of white and black marbles.
06 Jin: Because there are white and black marbles, how about dividing 300 by 2 and say 150.
07 Kim: There can be 299 white marbles, can be 298 white marbles, can be 297 white marbles, hence, I think we can say any number.

The teacher stopped Activity 1 and moved into Activity 2 because the students no longer arrived at other methods. The teacher then introduced Activity 2 to them. She showed the students a Korean story book and required them to design methods to predict how many times the Korean letter 'ga′ would appear in the book. First, they were given some time to find the methods.

08 Teacher: How we can predict that? Anyone have any idea?
09 Kim: We can count all the 'ga' characters.
10 Teacher: Is there any other way?
11 Choi: (On page 27) 'ga' appears twice on it. And, can we multiple the number of pages by two?
12 Teacher: Can we predict that 'ga' appears twice on it (on page 26), too?
13 Jin: Do you[the teacher] know the answer?
14 Lee: I guess she doesn't know, too.
15 Teacher: You can say the approximate number of 'ga' character, not the exact number.
16 Kim: Then, let's count together 'ga' on page 28, and multiple by half [of the number of pages].
17 Teacher: Anyone have any other idea?
18 Jin: We can suppose there is one 'ga' character per line, and count the number of lines.
19 Teacher: (On page 27), this line includes one 'ga', but this one, this one, this one, they do not include 'ga'.
20 Jin: Because we now predict, we can say there is one 'ga' per line… (equivocating)
21 Lee: It does not make sense.

The absence of a picture of surveying - investigating a sample instead of its population to obtain the information of the population - was a major obstacle when finding the methods in Activity 1 and 2. The students thought that they had to investigate the entire set (all of the marbles in the jar or all instances of 'ga' on all pages of the book) to predict it.

Activities' contributions

Contributions of Activity 3

The teacher stopped Activity 2 and moved on to Activity 3 because the students no longer arrived at other methods. The teacher introduced Activity 3 to them. The activity included a survey of popular foods for school meals. First, the students were given some time to design the methods.

22 Teacher: Is there anyone with any ideas?
23 Choi: We can give all of the students in our school questionnaire.
24 Lee: I think it is too difficult.
25 Teacher: Why do you think so?
26 Lee: We have too many students and it will take too long to ask all of them. Plus it will be too hard to summarize the data.
27 Kim: Can we choose a representative?
28 Teacher: What representatives?
29 Kim: To reduce the number of students asked… If there are too many students, it takes too long, and it will be hard. I think it is good to select representatives.
30 Teacher: Well, how many students are okay as representatives?
31 Kim: One student from every class.
32 Jin: But, each grade has a different number of classes and each class has a different number of students.
33 Teacher: Do you think that is a problem?
34 Jin: Yes, I do. I think each class has to have the same number of students to select one representative from each class.

35 Teacher: What we can do?

The students were given some time to think up new ideas.

36 Teacher: Do you have a new idea?
37 Kim: We can give a quarter of each grade a questionnaire.
38 Choi: We can ask the teachers which foods have been left the most and the least.

The teacher writes ideas provided by the students on the board.

39 Teacher: You gave three ideas. The first is to give all of the students in your school a questionnaire. The second is to select one student from each class. The last is to survey a quarter of each grade. Which do you think is the best idea?
40 Kim: The second one.
41 Jin: Me, too.
42 Choi: Me, too.
43 Lee: Me, too.
44 Teacher: Why do you think the second one is better than the first one?
45 Kim: It takes too long to ask all of the students, and it will be hard to summarize the data.
46 Teacher: Any other reason?
47 Students: (All the students agree on Kim's opinion)
48 Teacher: And then, why do you think the second one is better than the last one?
49 Kim: I want to change my answer to the last one.
50 Teacher: Why?
51 Kim: It is the same in the second one and the last one not to survey all of the students in our school, but we can gather more opinions, a greater variety of opinions in the last one than in the second one.
52 Choi: But selecting representatives of each class would be the same.
53 Teacher: But there is a difference in the number of representatives. There is one student in the second one but 7~8 students in the last one. Which is better?
54 Kim: The last one.
55 Lee: The last one.

Activity 3 was familiar to the students and could therefore stimulate their thinking related to surveying with a sample and with sampling. The students could form a picture of a survey with a sample instead of the entire population to obtain information about the population. They agreed on surveying a quarter of all students (a sample) instead of all students (the population) without trouble.

Contributions of Activity

Activity 1 was followed by Activity 3 again.

56 Teacher: What we can do if the first idea is applied here?
57 Kim: We can count the black and white marbles in all of the marbles.
58 Teacher: And then, what we can do if the second or last idea is applied here?
59 Kim: We can draw out a quarter of the marbles from the box and count the black and white marbles.
60 Teacher: And?
61 Kim: We need to multiple the numbers of black and white marbles by four.

The teacher drew out 75 marbles from the box in front of the students. They counted the black and white marbles among the 75 marbles. There were 52 black marbles and 23 white marble. They then predicted 208 black marbles and 92 white marbles in the box. There were 200 black marbles and 100 white marbles in the box.

62 Kim: We almost predicted the exact numbers of marbles.
63 Choi: I didn't know surveying a quarter of the marbles can make such an accurate prediction. It's good.
64 Lee: It [surveying a quarter of the marbles] is better than counting all of them.

Activity 1, in which the information about the population can be identified at once, provided the students an opportunity to recognize the power of a sample and to trust the information from a sample.

Contributions of Activity 2

Activity 2 was followed by Activity 1 again.

65 Teacher: What we can do if the first idea is applied here?
66 Students: We can count 'ga' on all of the pages.
67 Teacher: And, what can we do if the second or last idea is applied here?
68 Choi: All we have to do is to investigate a quarter of all the pages and multiple it by 4.
69 Teacher: The book has around 200 pages, so that we need to count 'ga' on 50 pages. Which pages should we count 'ga' on?
70 Choi: Let's count 'ga' [From page 1] to page 50.
71 Kim: No, we need to select the pages here and there.
72 Teacher: Here and there? How?
73 Kim: Randomly.
74 Teacher: At random? How?
75 Lee: We select pages 4, 8, and 12, … it will be good.

Although the students did not use the word *random* with a full understanding, we can consider that they recognized the importance of unbiased samples. Activity 2 provided them the opportunity to think of sampling methods.

CONCLUSION

From the results we can obtain two findings. One is that the absence of a picture of surveying - investigating a sample instead of its population to obtain the information of the population – is one of various obstacles in developing ideas related to samples and sampling. The other is that the participants can experience and notice different ideas regarding samples and sampling in different activities. In the Activity 3, they acknowledged that samples are useful for obtaining the information about populations. A population survey is difficult and is not overly useful. In the Activity 1, they recognized that samples cannot be identical to their population but that the information from a group of samples is similar to the information of the population. In the Activity 2, they devised some ideas about random sampling even though the ideas were immature.

REFERENCES

Australian Education Council. (1991). *A national statement on mathematics for Australian schools*. Carlton, Vic.: Author. Retrieved November 23, 2010 from http://www.eric.ed.gov/PDFS/ED428947.pdf

Batanero, C., Godino, J. D., Vallecillos, A., Green, D. R., & Holmes, P. (1994). Errors and difficulties in understanding elementary statistical concepts. *International Journal of Mathematics Education in Science and Technology, 25*(4), 527–545.

Department for Education. (1999). *Mathematics: The National Curriculum for England*. Wellington, New Zealand: Author. Retrieved November 23, 2010 from http://publications.education.gov.uk/ eOrderingDownload/QCA-99-460.pdf

Fischbein, E. & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal of Research in Mathematics Education, 28*(1), 96-105.

Garfield, J. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education, 10*(3). [Online: www.amstat.org/publications/jse/v10n3/garfield.html]

Jacobs, V. R. (1999). How do students think about statistical sampling before intruduction? *Mathematics in the Middle School, 5*, 240-263.

Lipson, K. (2002). The role of computer based technology in developing understanding of the sampling distribution. In B. Phillips (Ed.), *Proceedings of the 6th International Conference on Teaching Statistics*. [CD-ROM] Voorburg, The Netherlands: International Statistics Institute. [Online: http://www.stat.auckland.ac.nz/~iase/publications/1/6c1_lips.pdf]

Metz, K. E. (1999). Why sampling works or why it can't: Ideas of young children engaged in research of their own design. In F. Hitt & M. Santos (Eds.), *Proceedings of the 21st annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp.492-498). Cuernavaca, Mexico: PME.

Ministry of Education. (1992). *Mathematics in the New Zealand curriculum*. Wellington, New Zealand: Author. Retrieved November 23, 2010 from http://www.minedu.govt.nz/~/media/MinEdu/Files/ EducationSectors/Schools/MathematicsInTheNZCurriculum.pdf

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

Pfannkuch, M. (2008). *Building sampling concepts for statistical inference: A case study*, paper presented at the ICME 2008 TSG. Monterrey, Mexico. [Online: http://tsg.icme11.org/document/get/476]

Saldanha, L. & Thompson, P. (2002). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics, 51*, 257-270.

Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp.465-494). New York: MacMillan.

Rubin, A., Bruce, B., & Tenney, Y. (1991). Learning about sampling: trouble at the core of statistics. In D. Vere-Jones (ed.), *Proceedings of the Third International Conference on Teaching Statistics* (pp. 314-319). Voorburg, The Netherlands: International Statistical Institute.

Watson, J. M. (2004). Developing reasoning about samples. In D. Ben-Zvi and J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 277-294). Dordrecht, The Netherlands: Kluwer Academic Publishers.

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

Watson, J. M. & Moritz, J. B. (2000). Developing concepts of sampling. *Journal of Research in Mathematics Education, 31*(1), 44-70.