Types of Student Reasoning on Sampling Tasks

J. Michael Shaughnessy, Matt Ciancetta, & Dan Canada
Portland State University

As part of a research project on students' understanding of variability in statistics, 272 students, (84 middle school and 188 secondary school, grades 6 – 12) were surveyed on a series of tasks involving repeated sampling. Students' reasoning on the tasks predominanly fell into three types: additive, proportional, or distributional, depending on whether their explanations were driven by frequencies, by relative frequencies, or by both expected proportions and spreads. A high percentage of students' predominant form of reasoning was additive on these tasks. When secondary students were presented with a second series of sampling tasks involving a larger mixture and a larger sample size, they were more likely to predict extreme values than for the smaller mixture and sample size. In order for students to develop their intuition for what to expect in dichotomous sampling experiments, teachers and curriculum developers need to draw explicit *attention to the power of proportional reasoning in sampling tasks. Likewise, in order for students to develop their sense of expected variation in a sampling experiment, they need a lot of experience in predicting outcomes, and then comparing their predictions to actual data.*

This study builds on previous research on middle school students' understanding of variability in a repeated sampling environment conducted by Shaughnessy et al (2003), Watson et al (2003), Toruk and Watson (2000), and Reading & Shaughnessy (2000). The study adds to previous research by including a large number of subjects from grades 6 to 12, by suggesting a possible conceptual analysis of types of students' reasoning in repeated samples tasks, and by including tasks with several population sizes and sample sizes. This work is part of an ongoing research project to investigate students' conceptions of variability in a variety of contexts.

Subjects & Procedures

A series of questions involving a sampling context was administered in survey form to 272 middle school (N = 84) and secondary school (N = 188) students, mostly from a large metropolitan area in the northwestern part of the United States. The students were in ten classrooms from six schools—two urban, three suburban, and one rural–in two middle schools and four high schools. All six schools are participating in an ongoing research project on students' understanding of variability, with a teacher in each school serving as a consultant and co-researcher on the project. Students in all six schools had some previous experience with graphing data. Three of the four high schools and both middle schools use curriculum materials that include statistics investigations and probability experiments. Students in the other high school had no previous exposure to probability, and little to statistics. This multi year project includes survey tasks, interview tasks, and classroom teaching episodes that involve variability. In this paper we will concentrate mainly on a subset of the survey tasks that were given to all the students at the beginning of the project, prior to the teaching episode work of the project.

All 272 students were surveyed on a series of sampling tasks at the beginning of the project. The first series of tasks involved a mixture of 100 candies, 60 red and 40 yellow, which were thoroughly mixed. Handfuls of ten candies were to be pulled out, the number of reds would be recorded after each pull, and the candies would be put back in the mixture and remixed for the next pull of 10. The students were asked this series of questions:

1) How many reds would you expect to get in a handful (of 10 candies)? Why?
2) Would you expect to get that number of reds every time if you did it several times? Why?
3) What would surprise you? How many reds would surprise you in a handful of ten? Why would that surprise you?
4) What numbers of reds would you predict for six handfuls? (Each time candies replaced and remixed before pulling again) Why did you make those predictions?
5) Construct a graph of the results for the numbers of reds for 50 handfuls of ten candies.

A series of similar questions was given only to the secondary students using a mixture of 1000 candies, 600 red, 400 yellow, and sample size 100. Both the population proportion (60% red) and the relative sample size (10% of the population) were constant for the large and small mixtures.

Method

Students' responses to each question were categorized, and then coded on a scale (0, 1, 2, etc) with higher numbers indicating more student use of variation reasoning and/or proportional reasoning. The coding schemes for the items were developed iteratively over several runs by a team of three researchers using the responses from two classes. Subsequently each researcher independently scored every student on each item on the remaining classes. Initial inter-rater agreement percentages were 100%, 82%, 90%, 94%, and 97% for the items presented in this paper. Any disagreements were subsequently discussed and resolved, so that in the end all three researchers agreed on the final coding of each response.

Types of Reasoning

Students' responses and reasoning on these questions fell mainly into three broad categories: additive, proportional, or distributional reasoning. Additive responses tended to rely on absolute numbers or frequencies of reds in the original mixture, e.g. *"because there are more reds."* Proportional reasons fell into two subgroups. Some students' responses implicitly suggested that they used sample proportions or population proportions, or probabilities, or percents in their thinking, but they had difficulty putting their reasoning into words. ("*Most of them will be around 6, but I just can't explain why"* (implicit proportional reasoners*)*. Other students explicitly mentioned 'ratio of reds', 'percent of reds', 'probability of reds' in their reasoning, and connected it back to the original mixture (explicit proportional reasoners). Distributional reasons integrated both centers, and variation around those centers, into their reasoning on these tasks. A summary of students' responses to the first four questions listed above is presented in Tables 1 – 5, along with codes and code descriptors for each item. The question that asks students to graph the results of 50 samples of 10 will not be discussed in this paper due to space limitations, but we mention it so that readers are aware that students were also asked about larger numbers of repetitions.

Results on Sampling Tasks

1. Suppose you have a container with 100 candies in it. 60 are red, and 40 are yellow. The candies are all mixed up in the container. You pull out a handful of 10 candies.

How many reds do you expect to get? __________

Table 1. Responses to number of reds expected in one handful of ten.

Responses	Code	MS (N=84)	HS (N=188)	All (N=272)
Other than 6	0	15	25	40
Six reds	1	64	160	224
A range (i.e. 5-7)	2	5	3	8

Codes for this item 0 – Other than around 6
1 – Six Red
2 – About Six, or a range, e.g. 5 –7

Most of the students responded that they'd expect 6 red. Only 8 students out of 272 volunteered a range of possibilities for the number of reds that would be pulled, that is, only 8 spontaneously identified variability as an issue that might arise in this first task. Students focus right in on the expected value, as is oft the case when they are only asked about one trial. It is rather surprising that nearly 15% of the students wrote they expected to get something else than about 6.

2. Suppose you did this several times. Do you think this many reds would come out every time? Why do you think this?

Table 2. Responses to "Would you expect the same number every time?"

Response	Code	MS (N=84)	HS (N=188)	All (N=272)
Yes	0	13	47	60
No- Poor reason or additive reason	1	43	65	108
No - Acknowledged Variation around 6, implicit proportional reasoning	2	27	62	89
No- Explicit proportional reasoning; strong variation reasoning	3	1	13	14
No- Distributional reasoning	4	0	1	1

Codes for this item *Yes* codes: Usually coded 0; occasionally students wrote 'yes', but their reasoning indicated they knew things would vary. Such cases were coded according to the *No* code scheme below.

No codes: 1—no reason given; vague or nonsense reason; "could be anything" reasoning; additive reasoning such as "there are more reds"

2—Some implicit indication of variation, "around 6"—but no explicit information about the distribution or about proportional reasoning, e.g., *"won't be the same every time."; "probability is not exact every time"*

3—Explicit Reasoning using the ratio, average, percent, or chance of reds (60% reds, 6 : 4 ratio); or reasonable spread. Some clear indication was given of proportional reasoning about the distribution of outcomes.
4—Explicit use of both a reasonable spread, as well as a spread around the expected value—distributional reasoning

25% of the HS students agreed, *yes, it will be the same every time*. This is consistent with findings in previous research (Shaughnessy et al, 1999; Reading & Shaughnessy, 2000; Shaughnessy et. al, 2003). The influence of the probability teaching may interfere with students thinking about variability. "Six reds" is *supposed* to happen, theoretically, in the minds of many students, because that is what probability says. This type of thinking about probability, particularly among the high school students waffled during our extended interviews, as the tension between "the most likely individual outcome", and a "likely distributions for a set of repeated outcomes", became more evident when follow-up questions were possible. Only 15 students gave reasons that explicitly used proportions (explicit proportional or distributional reasons). Two-thirds of the students did not reason proportionally at all on this task. Many students relied on additive thinking, such as "there are more red" or on "anything can happen". This latter response is reminiscent of the outcome approach discussed by Konold et al (1993).

3. How many reds would surprise you? __________Why would that surprise you?

Table 3. Responses to "How many reds would surprise you in a handful of ten?"

Response	Code	MS (N=84)	HS (N=188)	All (N=272)
Any number from 4 to 8	0	11	39	50
0 –3, 9, 10; blank, or additive reasoning	1	59	98	157
0 –3, 9, 10; proportional or distributional reasoning	2	13	44	57
Mentioned both ends, and used proportional reasoning	3	1	7	8

Codes for this item 0 – 4, 5, 6, 7, 8
1 – 0,1,2,3, 9, 10 *"because there are more reds"*
2 – Same numbers plus adequate reason (Which means that they attend to proportions or features of the distribution: Average, Ratio, Spread, chance)
(+1) If they mention both ends of the distribution

Students who responded that a number from 4 to 8 reds would be surprising were coded zero on this question, because these outcomes account for most of the cumulative probability distribution, and they really aren't surprising outcomes. 80% of the students identified at least one surprising outcome, in the sense that it had a low probability of occurring. However, only about 25% could explain why those numbers were surprising by appealing to features of the distribution, such as the proportion of reds, or the spread of outcomes. Most of the responses that were coded 1 were students who just put down one or two numbers that would surprise them, such as 10 reds, or 1 red. Students tend to believe that the extreme outcomes (0, 1, 9, 10) will occur much more frequently than they actually do in practice, as verified later on by their predictions for repeated samples in the classroom teaching episode where we carried out the sampling with them. There is a long research history dating back to early work in cognitive psychology (Kahneman & Tversky, 1972) which indicates that people lack intuition for the shape of probability distributions in dichotomous sampling tasks. Students' responses on this item bear out that lack of intuition.

4a.Suppose that six of your classmates do this experiment, each of them pulling out 10 candies. (After each pull, the candies are put back and remixed).
What do you think is likely to occur for the numbers of red candies that each classmate would pull out? ______, ______, ______, ______, ______, ______
Why do you think this?

Table 4. Summary of responses for six pulls (of 10) from the 60 – 40 mixture.

Responses	Code	MS (N=84)	HS (N=188)	All (N=272)
Too much or too little variation (N, W, H, L)	0	24	93	117
Appropriate choices but additive or poor reasoning	1	51	57	108
Centers *or* Spreads, Proportional reasoning	2	9	32	41
Centers *and* Spreads, Distributional reasoning	3	0	6	6

Codes for this item 0 – Too much or too little variation,
e.g., W(ide)-range ≥8, N(arrow)-range ≤1,
H(igh)-all ≥ 6, L(ow)-all ≤ 6
1 - Appropriate range of choices, but inappropriate or additive reasoning
e.g., *"there are more reds";" they are all mixed up"*
2-- Using ratio or average or chance or spread—some indication of proportional reasoning
3- Explicitly using variation combined with centers (distributional reasoning)

The main purpose of this item was to gain some insight into what students would predict for the results of repeated samples. Often students are only asked "what would you expect to happen" for one trial in a probability experiment or a sampling situation. Responses such as 3,7,5,6,6,4 or 7,5,6,8,3,6, or 5,6,6,7,6,6 were coded as having a reasonable or appropriate spread or variation, while choices such as 6,6,6,6,6,6 (too narrow, N), 1,7,3,9,10,4 (too wide, W), 1,2,3,4,5,6 (too low, L), or 6, 7, 6, 8, 9, 8 (too high, H) were coded 0, as they had too much or too little spread, or didn't bracket the expected value, 6. Nearly *half* the HS students were coded N, W, H, or L. The HS students did much worse on this task than the middle school students. Table 5 shows the breakdown for the 117 students who received a score of 0 (H, L, W, N) when predicting the results of six repeated samples of size 10 drawn from the 60 – 40 mixture.

Table 5. Breakdown of the 0 codes assigned to the 60 – 40 mixture.

Responses	MS (N=24)	HS (N=93)	All (N=117)
H	4	4	8
L	5	26	31
N	7	24	31
W	4	11	15
N & H	2	5	7
N & L	0	3	3
Other	2	20	22

The HS students who were coded 0 on the repeated sampling task in the 60 – 40 mixture tended to predict low, or narrow. The narrow (N) predictions (e.g., 6,6,6,5,6,6) accounted for 25% of the zero codes and could be an influence of probability instruction, or just lack or exposure to statistics tasks involving variability. Another 25% of the responses were coded 0 because they were Low (L). These students may tend to think of 6 reds as an upper bound for the number of reds that one could get in a handful. Such thinking shows a complete lack of understanding of how sampling results are distributed around a center. Most of the (31) Low responses occurred among 9th graders in the school where students had little or no previous exposure to statistics or probability. The large number of low predictions in this sample of students contrasts with previous results where more students predicted high (H) than low (L) (Shaughnessy et al, 1999). That earlier pilot study was conducted with students primarily in grades 4 – 6, while this study was conducted with older students, grades 6 – 12. Perhaps younger students are more likely to be influenced by the "larger number of reds" in the mixture than older students, and thus predict higher. (Note: The "Other" category in Table 5 includes blank responses, or written word responses, such as red, red,…, red, written in for the six pulls).

A similar series of questions on a mixture of 1000 candies, 600 red and 400 yellow, was administered only to the 188 secondary students in the study (due to time constraints). The results of students' predictions for six pulls from this mixture are presented in Table 6.

Table 6. Summary of responses for six pulls (of 100) from the 600 – 400 mixture

Responses	Code	HS (N=188)
Too much or too little variation (N, W, H, L)	0	123
Appropriate choices but additive or poor reasoning	1	34
Centers *or* Spreads, Proportional reasoning	2	26
Centers *and* Spreads, Distributional reasoning	3	5

Codes for this item 0 – Too much or too little variation,
& W(ide) range ≥50, N(arrow) range ≤3, H(igh all≥60, L(ow) all≤60
1 - Appropriate range of choices, but inappropriate or additive reasoning, e.g., *"there are more reds";" they are all mixed up"*
2-- Using ratio or average or chance or spread—some indication of proportional reasoning
3- Explicitly using variation combined with centers (distributional reasoning)

Of the 188 secondary students, 65% did not have a good feel for what would be likely to occur in six pulls. The breakdown for the 123 students who received a score of 0 when predicting the results of six pulls from the 600 – 400 mixture was as follows: 30 L; 26 N; 21 W; 4 H; 5 W&L; 3 N&H; 1 N&L; 1 W&H; and 32 Other; (blank or words written in). The population proportion (60% reds) in this larger mixture was the same as for the 60 – 40 mixture, and the relative sample size was maintained at 10% of the population. However, performance was much worse on this task than on the 60 – 40 mixture, with two-thirds of the students making choices for the numbers of reds in their repeated samples that were W, N, H, or L. Students who were coded 0 on this task often predicted low, narrow, or wide. With a much bigger range of numbers, the students were more likely to predict wide for this sampling task than for the smaller mixture. This task provides further evidence of the tendency, also noted in the studies cited above, for students to predict too wide a range of outcomes, or to believe that outcomes with very low probabilities *will* occur.

Results & Discussion

- Only 8 of our 272 students spontaneously acknowledged the possibility of variation in a sampling situation when there was only one trial (Question 1). This type of question doesn't even raise the role of variability in sampling. Furthermore, it concentrates student thinking on centers, as opposed to spreads. We recommend against using such questions in isolation from other questions on sampling, such as our questions 2 – 5, because they mask variability.
- When asked if they will get the same result every time they sample, surprisingly 25% of our students said yes, they will. However, based on our experience with interviewing students using similar tasks, we believe that many of these students would qualify there thinking under further questioning, and say things like "that's what theory predicts, but you might not get that if you actually did it, even though your supposed to." There may be interference from past experiences with probability that distract from variability.
- Many of our students did not have a good sense for the results of a repeated sampling situation, particularly when a large sample size is drawn. Some students believe that a very wide range of possible outcomes will always occur in a dichotomous sampling task. Others predict a very narrow band of outcomes, while still others predict a range (too high or too low) that does not even bracket the population proportion (60% in this case).
- Our students tended to believe that extreme outcomes *will* occur. Many students wrote that only the most extreme outcomes (0 or 10; 0 or 100) would surprise them. Only 8 of 272 students spontaneously identified surprising outcomes in both tails of the distribution. Again, during probing in interviews, we have found that students will mention extreme outcomes at both ends as surprising, but only when specifically asked.
- Our students tended *not* to use the potential power of proportional reasoning in their explanations for their responses. They relied more on additive or frequency types of arguments than on proportions or relative frequencies in their responses. The percentage of students who used proportional (or distributional) reasoning on questions 2, 3, 4a, and 4b, were respectively 38%, 24%, 17%, and 16%, This

suggests that students do not evoke the connections that proportions have to sampling situations, or that they are weak proportional reasoners in general.

We believe that our students are not very that different than your students, since our students come from a variety of school settings, a variety of socio-economic situations, and a variety of teaching and curriculum situations. Proportional reasoning is the cornerstone of statistical inference. In order for students to develop their intuition for what to expect in dichotomous sampling situations, we strongly recommend that teachers and curriculum developers provide many more opportunities to enhance students' proportional reasoning skills when working in a sampling environment. Furthermore, to improve students' feel for the expected variability in a sampling situation, students need considerable hand's on experience in first predicting the results of samples, and then drawing actual samples, graphing the results, comparing their predictions to the actual data, and discussing observed variability in the distribution. A forthcoming article (Shaughnessy & Watson, in press) provides several such opportunities for teachers to enhance students' proportional reasoning skills in statistical settings. The power of proportional reasoning in statistical situations needs to be identified much more explicitly in order for our students to evoke the connections of proportional thinking to statistical settings.

References

Kahneman, D, & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*, 430-454.

Konold, C., Pollatsek, A., Well, A., Lohmeier, J. & Lipson, A. (1993). Inconsistencies in students' reasoning about probability. *Journal for Research in Mathematics Education, 24*, 392-414.

Reading, C. & Shaughnessy, J. M. (2000) Student perceptions of variation in a sampling situation. In T. Hakahar & M. Koyama (Eds.), *Proceedings of the 24th Conference of the International Group for the Psychology of Mathematics Education* (vol. 4, pp. 89 – 96) Hiroshima, Japan.

Shaughnessy, J. M., Canada, D., & Ciancetta, M. (2003). Middle School Students' Thinking About Variability in Repeated Trials: A Cross -Task Comparison. In N. Pateman, B. Dougherty, & J. Zillox, (Eds.), *Proceedings of 27th Conference of the International Group for the Psychology of Mathematics Education* (vol. 4, pp. 159 - 166) Honolulu, USA.

Shaughnessy, J. M., Watson, J., Moritz, J., & Reading, C. (1999). School mathematics students' Acknowledgement of statistical variation. NCTM Research Pre-session Symposium: *There's More to Life than Centers.* Paper Presented at the 77th Annual NCTM Conference, S. F., CA.

Torok, R. & Watson, J. M. (2000). Development of the concept of statistical variation: An exploratory study. *Mathematics Education Research Journal,* 12, 147-169.

Watson, J.M., Kelly, B.A., Callingham, R.A., & Shaughnessy, J.M. (2003). The measurement of school students' understanding of statistical variation. *International Journal of Mathematical Education in Science and Technology, 34,* 1 – 29.

Watson, J. & Shaughnessy, J. M. (in press). Proportional Reasoning: Lessons from Research in Data and Chance. *Mathematics Teaching in the Middle School.*