STUDENTS' SCHEME-BASED CONCEPTIONS OF SAMPLING AND ITS RELATIONSHIP TO STATISTICAL INFERENCE

Luis Saldanha
Vanderbilt University
luis.a.saldanha@vanderbilt.edu

Patrick Thompson
Vanderbilt University
pat.thompson@vanderbilt.edu

We distinguish two conceptions of sample and sampling that emerged in the context of a teaching experiment conducted in a high school statistics class. In one conception "sample as a quasi-proportional, small-scale version of the population" is the encompassing image. This conception entails images of repeating the sampling process and an image of variability among its outcomes that supports reasoning about distributions. In contrast, a sample may be viewed simply as "a subset of a population"- an encompassing image devoid of repeated sampling, and of ideas of variability that extend to distribution. We argue that the former conception is a powerful one to target for instruction.

Background

On the basis of empirical evidence Kahneman and Tversky (1972) hypothesized that people often base judgments of the probability that a sample will occur on the degree to which they think the sample "(i) is similar in essential characteristics to its parent population; and (ii) reflects the salient features of the process by which it is generated" (ibid., p. 430). This hypothesis suggests that Kahneman and Tversky's subjects focused their attention on *individual* samples. In later research, Kahneman and Tversky (1982) conjectured that people, indeed, tend to take a *singular* rather than a *distributional* perspective when making judgments under uncertainty. In the former, one focuses on the causal system that produced the particular outcome and assesses probabilities "by the propensities of the particular case at hand" (ibid., p. 517). In contrast, the distributional perspective relates the case at hand to a sampling schema and views an individual case as "an instance of a class of similar cases, for which relative frequencies of outcomes are known or can be estimated" (ibid., p. 518).

Konold (1989) found strong empirical support for Kahneman and Tversky's (1982) conjecture. He presented compelling evidence that people, when asked questions that are ostensibly about probability, instead think they are being asked to predict *with certainty* the outcome of an *individual* trial of an experiment. Konold (ibid.) characterized this orientation, which he referred to as the *outcome approach*, as entailing a tendency to base predictions of uncertain outcomes on causal explanations instead of on information obtained from repeating an experiment.

Sedlmeier and Gigerenzer (1997) analyzed several decades of research on understanding the effects of sample size in statistical prediction. They argued compellingly that subjects across a diverse spectrum of studies who incorrectly answered tasks involving a distribution of sample statistics may have interpreted task situations and questions as being about individual samples.

Recent instructional studies (delMas, 1999; Sedlmeier, 1999) indicated that engagement in carefully designed instructional activities using computer simulations of drawing many samples can help orient students' attention to collections of sample statistics when making judgments involving samples. However, analyses in these studies did not focus on characterizing students' evolving conceptions and imagery in relation to their engagement in instruction.

Despite the centrality of variability in statistics, students' understanding of sampling variability and our comprehension of variability's role as a central organizing idea in statistics instruction has received little research attention (Shaughnessy et al., 1999). Rubin et al. (1991) proposed that a coherent understanding of sampling and inference entails integrating ideas of sample representativeness and sampling variability to reason about distributions. Images of the re-sampling process, however, were not at the foreground of their conceptual analysis. Other conceptual analyses of sampling (Schwartz et al., 1998; Watson and Moritz, 2000) characterized the relationship between population and a randomly selected subset of it in a way that did not entail images of the repeatability of the sampling process nor of the variability that we can expect among sample outcomes.

In sum, substantial evidence from research on understanding samples and sampling suggests that students tend to focus on individual samples and statistical summaries of them instead of on how collection*s* of sample statistics are distributed. Furthermore, students may tend to predict a sample's outcome on the basis of causal analyses instead of statistical patterns in a collection of sample outcomes. These orientations are problematic for learning statistical inference because they disable students from considering the relative unusualness of a sampling process' outcome. Finally, sampling has not been characterized in the literature as an interrelated scheme of ideas entailing repeated random selection, variability, and distribution.

Purpose and Methods

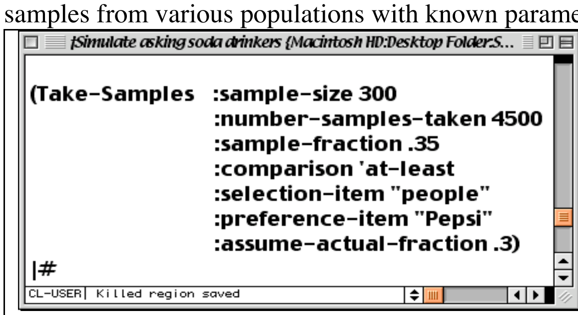
This study investigated the development of students' thinking as they participated in instruction designed to support their conceiving sampling as a scheme of interrelated ideas including repeated random selection, variability among sample statistics, and distribution.

Twenty-seven 11th- and 12th-grade students, enrolled in a non-AP semester-long statistics course, participated in a 9-session whole-class teaching experiment (TE) addressing ideas of sample, sampling distributions, and margins of error. Our aim was to develop epistemological analyses of these ideas (Glasersfeld, 1995; Steffe & Thompson, 2000; Thompson & Saldanha, 2000) – ways of thinking about them that are schematic, imagistic, and dynamic – and hypotheses about their development in relation to students' engagement in classroom instruction.

Three research team members were present in the classroom during all lessons: one author designed and conducted the instruction; the other author observed the instructional sessions and took field notes; a third member operated the video cameras. Students' understandings were investigated in three ways: by tracing their participation in classroom discussions (all instruction was videotaped), by examining their written work, and by conducting post-experiment individual interviews.

Instruction stressed two overarching and related themes: 1) the random selection process can be repeated, and 2) judgments about sampling outcomes can be made on the basis of relative frequency patterns that emerge in collections of outcomes of similar samples.[i] These themes were intended to support students' developing a distributional interpretation of sampling and likelihood. Though an a priori outline of the intended teaching and learning trajectories (Simon, 1995) guided the progress of the teaching experiment, the research team made on-line adjustments to instruction according to what they perceived as important issues that arose for students in each session.

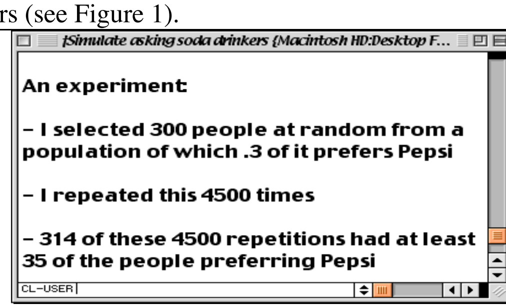
The teaching experiment unfolded in three interrelated phases: it began with directed discussions centered on news reports that mentioned data about sampled populations and news reports about populations per se (raising the issue of sampling variability). The experiment then progressed to questions of "what fraction of the time would you expect results like these?" This entailed having students employ, describe the operation of, and explain the results of computer simulations of taking large numbers of samples from various populations with known parameters (see Figure 1).



Explain what each number stands for in the command we have been using to instruct the computer to simulate drawing random samples from a population of soda drinkers.

Explain what information we will get after having run the simulation (with the values provided above).

What result do you expect the simulation will produce (with these values provided above)? Please justify your answer.



Interpret the simulation's output above.
How does your prediction compare to the result produced by the simulation?

Are they significantly different?
Are you surprised by this difference?
What might account for the difference?

What fraction of the time would you expect results like these?

Figure 1. Part of an instructional activity designed to help students make sense of computer simulations of drawing many random samples from a population. Simulation input (left) and output (right) windows were displayed in the classroom and the instructor posed questions designed to orchestrate reflective discussions about the simulations.

The experiment ended by examining simulation results systematically, with the aim that students see that distributions of sample proportions are largely unaffected by underlying population proportions (see Figure 2), but are affected in important ways by sample size.

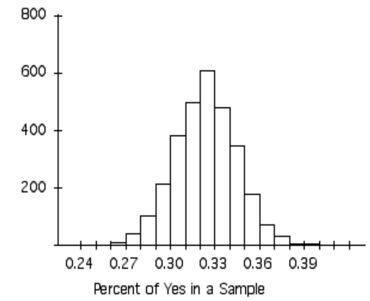
Here is a table similar to one we filled out in an in-class activity:

Percent of Yes in Population	Number of People in a Sample	Number of Samples Drawn	% of Sample Percents within 1 Percentage Point of Population %	% of Sample Percents within 2 Percentage Points of Population %	% of Sample Percents within 3 Percentage Points of Population %	% of Sample Percents within 4 Percentage Points of Population %
65%	500	2500	36.7%	64.5%	84.8%	91.5%
32%	500	3000	37.1%	65.8%	83.9%	91.1%
57%	500	6800	36.2%	64.9%	84.2%	91.3
60%	500	5500	36.1%	65.2%	84.3%	91.4%

a) Explain what "% of Sample Percents within 2 Percentage Points of Population %" means.

b) What do you infer from this table that is true about the process of taking a random sample of 500 people?

c) What does this suggest about a relationship between sample size and the percent of samples within certain percentage points of a population percent?



Levels	Counts
0.24000000	3
0.25000000	2
0.26000000	9
0.27000000	42
0.28000000	103
0.29000000	214
0.30000000	386
0.31000000	502
0.32000000	612
0.33000000	478
0.34000000	348
0.35000000	179
0.36000000	76
0.37000000	33
0.38000000	7
0.39000000	6

3000 Samples of 500 each taken from a population having .32 of it believing "yes"

Figure 2. Part of an instructional activity designed to structure students' investigation of the relationship between sampling distributions and underlying population proportions. Students filled out the table on the left by organizing information (like that shown on the right) generated by computer simulations of drawing many random samples from populations with given proportions.

Results and Discussion

In this report we move toward elaborating an important distinction between two conceptions of samples and sampling that emerged in the teaching experiment. Our analyses revealed that some students – generally those who performed better on the instructional activities and those who were able to hold coherent discourse about the mathematical ideas highlighted in instruction – had developed a multi-tiered scheme of conceptual operations centered around the images of repeatedly sampling from a population, recording a statistic, and tracking the accumulation of statistics as they distribute themselves along a range of possibilities. These images and operations were tightly aligned with those promoted in classroom instructional tasks and discussions. As such, we conjecture that these students' engagement in the instructional activities played an important role in their developing such a scheme. For instance, we had students practice imagining and describing a coordinated multi-level process that gives rise to sampling distributions (and to the simulations' results):

Level 1: Randomly select items to accumulate a sample of a given size from a population. Record a sample statistic of interest.

Level 2: Repeat Level 1 process a large number of times and accumulate a collection of statistics.

Level 3: Partition the collection in Level 2 to determine what proportion of statistics lie beyond (below) a given threshold value.

In classroom discussions the instructor employed a metaphor designed to help students distinguish and coordinate these different levels. The metaphor entails imagining a collected sample of dichotomous opinions ("yes" or "no") in Level 1 as a box containing "1"s (for "yes") and "0"s (for "no"). It then entails labeling the box with a "1" (or a "0") if the proportion of its contents is greater (or less) than a given threshold value. In this metaphor, what accumulates in Level 2 is a collection of "1"s and "0"s (or boxes/slips of paper labeled "1" or "0"), each of which represents a sample whose statistic is greater (less) than the threshold value. At Level 3, the metaphor entails calculating the percent of the collection of "1"s and "0"s in Level 2 that are "1" or that are "0", depending on the required comparison.

The following excerpt illustrates one student's coherent image of the multi-tiered sampling process, the development of which appeared to have been facilitated by his use of this metaphor. We take this student's coherent image as an expression of the stable scheme of conceptual operations characterized above. In the excerpt, the student (D) interpreted a sampling simulation's command and the result of running it as he viewed familiar simulation windows on a computer screen (see Figure 1)[ii]:

D: Ok. It's asking...the question is...like "do you like Garth Brooks". You're gonna go out and ask 30 people, it's gonna ask 30 people 4500 times if they like Garth Brooks. The uh...(talks to himself) what's this? let's see...the actual...like the amount of people who actually like Garth Brooks are...or 3 out of 10 people actually prefer like Garth Brooks' music. And uh...for the 30...when you go out and take one sample of 30 people, the cut off fraction means that if you're gonna count, you're gonna count that sample, if like 37% of the 30 people preferred Garth Brooks. And then it's going to tally up how many of the samples had 37% people that preferred Garth Brooks. So like the answer would be I don't know, like whatever, 2000 out of 4500 samples had at least 37% of people preferring Garth Brook.

[…]

I: How was it that you thought about it that allowed you to keep things straight? […]

D: I just thought of it like ... I don't know, I sort of thought of it like how you were saying. Like...if the like 1s and the 0s if you ask 30 uh if like 10 of them say they like Garth Brooks-- or for every person who likes Garth Brooks you put a 1 down, if they don't you put a zero. You do that 30 times and you're gonna get like I don't know, 15 ones and 15 zeros you add up, you add them up. Then it says the cutoff fraction for each sample is 37% so you have like at least 37% of the...like those or...30-- if you add it up and divided it by the 30 and it's at least 37% then you have like another pile of like little papers and you put a one on like the big, the big one for the sample or a zero if it's less than-- if the whole sample is less than 37%. The 1s and 0s I don't know...you said something about like...that sort of helped.

A significant feature of student D's thinking was his ability to clearly distinguish different levels of the resampling processes — never confounding the number of people in a sample with the number of samples taken — while coordinating the various levels into a structured whole. Additionally, and relatedly, student D interpreted the result of the simulation as an amount (percentage) of *sample proportions*, thus suggesting that he understood that the multi-level process generated a *collection* of sample proportions.[iii]

Student D's coherent image contrasts sharply with that of many poorer-performing students who persistently confounded numbers of people in a sample with numbers of samples drawn. The following interview excerpt illustrates one such student's (M) difficulties in the context of explaining similar computer simulations:

Segment 1

I: Ok, Suppose that, here's what I'm gonna do, uhh instead of 4500 samples I'm gonna take uhh, 1000 samples. Everything's gonna stay the same – sample size is 30, population fraction is 3/10ths, but now were' just taking 1000 samples. What would you expect the results to be?

[…]

M: Uhh, somewhere around like (short silence), hmm around like 25-30% of those 1000 samples.

I: Why 25-30%?

M: Because it's uhh…easier to uhh, I mean

I: What are you basing that judgment on?

M: Uhh, the actual population percentage, of 30

I: Ok, so you figure it'll be about 30%, 25 to 30, because the population fraction is 30%?

M: Yeah, somewhere close to that.

[…]

Segment 2

I: Alright (runs simulation, result displayed on output screen is "189 of these 1000 repetitions …")

M: 2/10ths, 20%. Hmm, it's still a little less

I: So it's a little less than 20%, right?

M: Hmm hmm, huh (seems surprised)

[…]

Segment 3

I: Alright. Suppose that now we, let's do this, let's make 2500 samples (changes parameter value in command window). What fraction of those samples, I mean what result would you now expect, for the number of samples that we're going to get that exceed 37% preferring Garth Brooks?

M: About 1/5 of those.

[…]

I: Now, before you would have said "well, 3/10ths of the 2500 samples, the 2500 repetitions"

M: Hmm hmm

I: Do you still sort of lean that way, that you should get around 3/10ths of the --?

M: I think it should, but I don't understand why it's not, why it keeps coming out with 1/5th rather than 1/3rd.

I: Alright, what is that "3/10ths" 3/10ths of?

M: Uhh, hmm 3/10ths of the entire population

I: Alright, and those are people, right?

M: Hmm hmm (nods)

I: Now, if you took 3/10ths of the 2500 repetitions you're taking 3/10ths of what?

M: Of the uhh…people sampled (chuckles)

I: No, 3/10ths of the samples.

M: Oh. Hmm hmm

Segment 1 of the excerpt suggests that student M expected the simulation to produce an amount (number) of samples and that he expected the percentage of that amount to hover around the sampled population percent (30%). Segment 2 illustrates his surprise at finding the actual percent being 20% of the 1000 samples generated. In segment 3 the student anticipated the same (20%) result for a simulation involving a larger number of samples, but he did not understand why this should be so because his conviction was that the simulation should produce a numerical value close to the sampled population percentage. The remainder of the segment reveals that student M had been interpreting the simulation's result as a percentage of people sampled rather than as a percentage of samples.

During such instructional activities most students experienced great difficulty conceiving the re-sampling process in terms of distinct levels. They would often unwittingly shift from speaking and thinking of a number of people in a sample to a number of samples selected. Their control of the coordination between the various levels of imagery was unstable; from one moment to the next their image of a number of samples (of people) seemed to easily dissolve into an image of a total number of people. These difficulties led many students to misinterpret a simulation's result as being about a percent of people rather than about a percent of sample proportions. This muddling of the different levels of the resampling process, in turn, obstructed their ability to imagine how sample proportions might distribute themselves around the underlying population proportion.

A salient consequence of these students' difficulty in imagining a sampling distribution was their tendency to judge a sample's representativeness only in relation to the underlying population proportion. Their image of sampling did not entail a sense of variability that extended to ideas of distribution: they understood that sample statistics vary, but only to the extent that if we were to draw more samples and compute statistics from them, those statistics would differ from the ones for the samples already drawn. Thus, judgments of a particular sampling outcome's unusualness were based largely on how they thought the outcome compared to the underlying population parameter per se, instead of on how it might compare to the way similar sample statistics were clustered around the parameter.

On the basis of such characteristics, we conjecture that these students' encompassing image of sample was *additive* — that is, in these instructional settings they tended to view a sample simply as a subset of a population and to view multiple samples as multiple subsets.

A contrasting image of sample is suggested in the following excerpt of student D explaining the purpose of simulating resampling:

D: If like...if you represent-- if you give it like the split of the population and then you run it through the how-- number of samples or whatever it'll give you the same results as if-- because in real life the population like of America actually has a split on whatever, on Pepsi, so it'll give you the same results as if you actually went out, did a survey with people of that split.

I: Ok, now. What do you mean by "same results"? On any particular survey at all--you'll get exactly what it--?

D: No, no. Each sample won't be the same but it's a...it'd be...could be close, closer...

I: What's the "it" that would be close?

D: If you get...if you take a sample...then the uh...the number of like whatever, the number of "yes"s would be close to the actual population split of what it should be.

I: Are you guaranteed that?

D: You're not guaranteed, but if you do it enough times you can say it's within like...1 or 2% of error depending upon uh how many times-- I think-- how many times you did it.

The resemblance between sample and population was clearly foremost in student D's mind, but his image was of a *fuzzy* resemblance bound up with ideas of variability and proto-distributional images of a collection of sample proportions. He did not expect a sample to be an exact replica of the sampled population, instead he anticipated that in repeating the sampling process many sample proportions would be "more or less" close to the population proportion. Moreover, student D's confidence in a sample's representativeness was based on this anticipated image of how a collection of similar sample proportions might be distributed around the population proportion.

We put that student D's description is consistent with his having conceived a sample as a quasi-proportional mini version of the sampled population, where the "quasi-proportionality" image comes from anticipating a bounded variety of outcomes, were one to repeat the sampling process.

It is often useful to refer to a germinating idea with suggestive terminology; we call this image of sample a *multiplicative conception of sample* (MCS) because its constitution entails conceptual operations of multiplicative reasoning. An elaboration of multiplicative reasoning (Harel & Confrey, 1994) is beyond the scope of this paper. For the present discussion we draw on Inhelder and Piaget's (1964) broad characterization of multiplicative reasoning as conceiving an object (quantity) as *simultaneously* composed of multiple attributes (quantities). For instance, conceiving a proportion involves multiplicative reasoning when it entails comparing two quantities in such a way as to think of the measure of one in terms of the measure of the other (Thompson & Saldanha, in press). An example is when one thinks of percentage as quantifying a part of a whole *in terms of the whole*. This conception entails keeping both the part and the whole simultaneously in mind and the ability to reciprocally relate and express one in terms of the other. This is different from thinking of measuring a subpart of a whole only in absolute terms.

We hypothesize that MCS entails multiplicative operations on several levels: on one level it entails conceiving a relationship of proportionality between a sample and a population. On another level, imagining the emergence of a *proto*-distribution of sample statistics entails structuring statistics as subclusters of the range of an entire collection of statistics. This involves fractional reasoning. Finally, a mature and well articulated image of distribution supports quantifying the expectation of a particular kind of sampling outcome and thus quantifying one's confidence in a sampling outcome's representativeness. This entails the operation of juxtaposing the individual sample result against an aggregate of similar sample results to compare the one against the many – an image of simultaneity that is central to multiplicative reasoning.

Conclusion

Though our elaboration of these two images of samples and sampling is empirically grounded, our point in presenting it is not to imply that students in our experiment fell into one or the other camp. Rather, our point is to highlight two significantly different conceptions and images of samples and sampling — perhaps exemplary of extremes in a continuum of students' conceptions — that provide insight into what may be more or less powerful conceptions to target for instruction.

From our perspective, there are two reasons why the distinction between the additive and multiplicative conceptions of sample is significant. First, in contrast to the additive conception, MCS entails a rich network of interrelated images that supports a deep understanding of statistical inference. In practice, statistical inferences about a population are typically made on the basis of information obtained from a *single* sample randomly drawn from the population. This practice is common among statisticians despite expectations of variability among sampling outcomes. In statistics instruction, however, it is uncommon to help students conceive of samples and sampling in ways that support their developing coherent understandings of *why* statisticians have confidence in this practice. We claim that MCS empowers students to understand the *why* by orienting them to relate individual sample outcomes to distributions of a class of similar outcomes. In the same way, MCS enables students to consider a sampling outcome's relative unusualness. As such, we propose that MCS characterizes a powerful "target" conception that can guide efforts to design instructional activities and student engagements intended to support their developing a deep understanding of sampling and inference.

The second reason why we consider the distinction between these two conceptions of samples to be significant is that few of our students developed MCS. Instead, most students seemed to tend toward an additive image of sample. To us, this state of affairs suggests that developing MCS is non-trivial. The reasons for students' difficulties in this regard are currently unclear to us. However, one plausible hypothesis grounded in our data is that for many students the simulation and sampling distribution activities were of such a complexity so as to essentially overshadow ideas of sampling variability highlighted in the first phase of the teaching experiment. In a subsequent teaching experiment (Saldanha & Thompson, 2001) we took this hypothesis seriously and engaged students in instructional activities designed to support their developing a MCS.

References

delMas, R. C., Garfield, J., & Chance, B. L. (1999). *Exploring the role of computer simulations in developing understanding of sampling distributions*. Paper presented at the American Educational Research Association, Montreal.

Glasersfeld, E. v. (1995). *Radical constructivism: A way of knowing and learning*. London: Falmer Press.

Harel, G., & Confrey, J. (Eds.). (1994). *The development of multiplicative reasoning in the learning of mathematics*. Albany, NY: SUNY Press.
Inhelder, B., & Piaget, J. (1964). *The early growth of logic in the child: classification and seriation*. New York: W. W. Norton & Company Inc.
Kahneman, D., & Tversky, A. (1982). Variants of uncertainty. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 509-521). New York: Cambridge University press.
Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgement of representativeness. *Cognitive Psychology*(3), 430-454.
Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction, 6*(1), 59-98.
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* (Vol. 1, pp. 314-319). Dunedin, New Zealand: ISI Publications in Statistical Education.
Saldanha, L. A. & Thompson, P. W. (2001). Students' reasoning about sampling distributions and statistical inference. In R. Speiser & C. Maher (Eds.), *Proceedings of the Twenty Third Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol.1, pp. 449-454). Snowbird, Utah: ERIC Clearinghouse for Science, Mathematics, and Environmental Education, Columbus, OH.
Schwartz, D. L., Goldman, S. R., Vye, N. J., & Barron, B. J. (1998). Aligning everyday and mathematical reasoning: The case of sampling assumptions. In S. P. Lajoie (Ed.), *Reflections on statistics: learning, teaching, and assessment in grades K-12* (pp. 233-273). Mahwah, NJ: Lawrence Erlbaum Associates.
Sedlmeier, P. (1999). *Improving statistical reasoning: Theoretical models and practical implications.* Mahwah, NJ: Lawrence Erlbaum.
Sedlmeier, P., & Gigerenzer, G. (1997). Intuitions about sample size: The empirical law of large numbers. *Journal of Behavioral Decision Making, 10*, 33-51.
Shaugnessy, J. M., Watson, J., Moritz, J., & Reading, C. (1999). *School students' acknowledgment of statistical variation*. Paper presented at the Research Presession Symposium of the 77th Annual NCTM Conference, San Francisco, CA.
Simon, M. A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education, 26*(2), 114-145.
Steffe, L. P., & Thompson, P. W. (2000). Teaching experiment methodology: Underlying principles and essential elements. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 267-306). Mahwah, NJ: Lawrence Erlbaum Associates. Inc.
Thompson, P. W., & Saldanha, L. A. (in press). Fractions and multiplicative reasoning. In J. Kilpatrick & G. Martin (Eds.), *Research companion to the Standards 2000*. Reston, VA: NCTM.
Thompson, P. W., & Saldanha, L. A. (2000). Epistemological analyses of mathematical ideas: A research methodology. In M. L. Fernandez (Ed.), *Proceedings of the Twenty Second Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 403-408). Tucson, AZ: ERIC Clearinghouse for Science, Mathematics, and Environmental Education, Columbus, OH.
Watson, J. M., & Moritz, J. B. (2000). Developing concepts of sampling. *Journal for Research in Mathematics Education, 31*(1), 44-70.

Research reported in this paper was supported by National Science Foundation Grant No. REC-9811879. Any conclusions or recommendations stated here are those of the authors and do not necessarily reflect official positions of NSF

[i] Similar samples share a common size, selection method, and parent population. Furthermore, they are selected to obtain information about a common population characteristic.

[ii] The simulation was of sampling people's preference for a particular musician from a hypothetical population having a known proportion of it preferring the musician.

[iii] We note that student D's prediction of the simulation result was highly inaccurate in this excerpt. Shortly thereafter, however, he quickly revised his prediction with a highly accurate one and continued to make such accurate predictions throughout the rest of the interview. We thus believe that his initial prediction was not an indication of a poor sense of how the sample proportions were distributed, rather it was merely the result of his focus, in the moment, on explaining *how* the simulation worked and *what* it generated.