Preliminary Research Report

From Beans to Polls: Does Understanding of Statistical Inference Within a Known Population Context Transfer to an Unknown Population Context?

Jennifer Noll, Portland State University
Sonya Redmond, Portland State University
Jason Dolor, Portland State University

Abstract: In today's society, informed citizenship requires at least an informal understanding of statistical inference. One strategy to promote such understanding is to develop students' knowledge of sampling distributions through simulation of repeated sampling from a population with a known parameter. An implicit premise in such instructional strategies is that students will be able to transfer their knowledge of sampling distributions created from a known population to "real-world" contexts such as public opinion polls. This preliminary report presents evidence suggesting that such transfer is neither immediate nor trivial. We will present case studies of two students enrolled in a statistics for teachers course illustrating the ways in which these two students, who displayed a robust knowledge of sampling distributions, applied this knowledge to polling scenarios.

Key Words: Statistical Inference, Sampling Distributions, Margin of error, Confidence

INTRODUCTION

An informal knowledge of statistical inference, including measures of statistical confidence and margin of error, is a vital skill for educated adults living in a global data based society. Popular media contains published results of opinion polls (often including confidence level and margin of error information), medical experiments, and advertising claims, all of which require readers to understand statistical arguments in order to participate intelligently in a democratic society and make good decisions about issues that effect their lives.

In the last decade and a half there have been significant increases in enrollments in high school AP statistics courses, as well as college level introductory statistics courses (Luzter, Rodi, Kirkman, & Maxwell, 2005), and this is likely in response to the need for a more statistically literate populace in an increasingly data driven society. Unfortunately much of the research (e.g., Bakker, & Gravemeijer, 2004; Garfield, delMas, & Chance, 2007; Heid et al., 2005; Konold, 1989; Liu, & Thompson, 2005; Watson, J. M. and J. B. Moritz, 2000) shows that students and their teachers struggle to understand introductory statistics topics. In addition, given that most introductory statistics courses focus on methods and procedures and do not incorporate examples of how statistics are used in the media, Gal (2003) argues that students are not likely to experience success in applying these methods and procedures to every day contexts. Thus, educational research must address the following questions: (1) what should be the goals of introductory statistics courses in order to prepare students to make informed data-based decisions, and (2) how can we best prepare teachers of statistics at effectively developing their students' statistical thinking skills? This preliminary research report begins to address these important questions through an investigation of pre-service teachers' statistical content knowledge. In particular, this report addresses pre-service high school teachers' knowledge of the conceptual underpinnings of statistical confidence (a key topic of introductory statistics courses) when applied to political polling scenarios (a statistical literacy application) via two mini-cases studies.

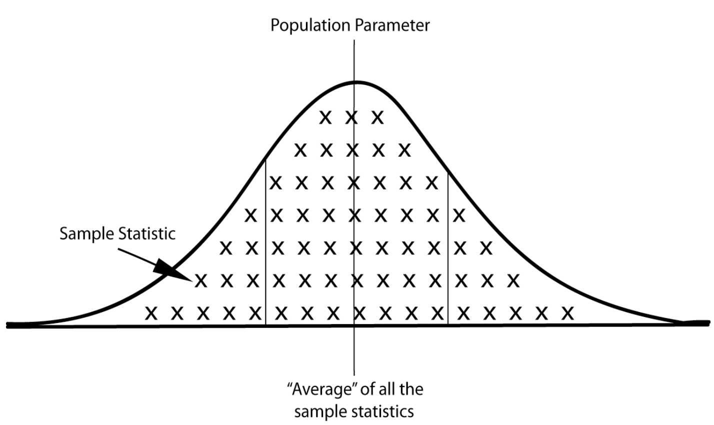
BACKGROUND

Statistical inference is a key topic of introductory statistics curricula. Thus, it is extremely important that teachers of statistics deeply understand statistical inference and the concepts that form the basis for this significant topic if they are to be effective in developing their students' statistical thinking skills. Many researchers (e.g., Chance, delMas, & Garfield, 2004; Garfield & Ben-Zvi, 2008; Saldanha & Thompson, 2007) have noted that an understanding of sampling distributions forms the basis for a robust informal understanding of statistical inference. These statistics educators argue that one method for developing students' understanding of informal statistical inference skills is through the concept of repeated sampling in which comparisons are made between samples and sampling distributions. An instructional strategy focused on sampling distributions attends to the underlying conceptual foundations of statistical inference. The subsection that follows provides an analysis of two concept definition images (in the manner of Tall & Vinner, 1981) underpinning a conceptual understanding of the level of confidence in statistics – distribution image and interval image. The final subsection provides background information on a common misconception of level of confidence.

A conceptual analysis of level of confidence in statistics

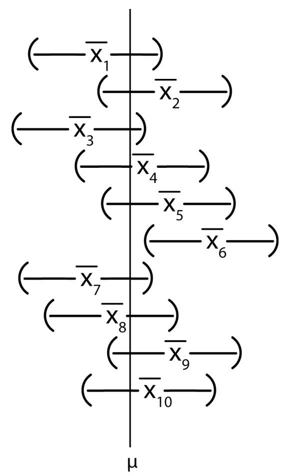
Defining the probability of a random event as a long-term relative frequency of outcomes, leads to two views for conceptualizing statistical confidence – a *distribution image* and an *interval image*. The distribution image entails constructing a distribution of sample statistics through repeated sampling. Empirical distributions of sample statistics can be approximated by a normal distribution and the center of the distribution is a good approximation for the population parameter. Figure 1 shows a visual representation of the distribution image. In the distribution image, the level of statistical confidence is defined as the percentage of sample statistics within a given distance – namely, margin of error – from the population parameter.

Figure 1. Distribution Image of Confidence.



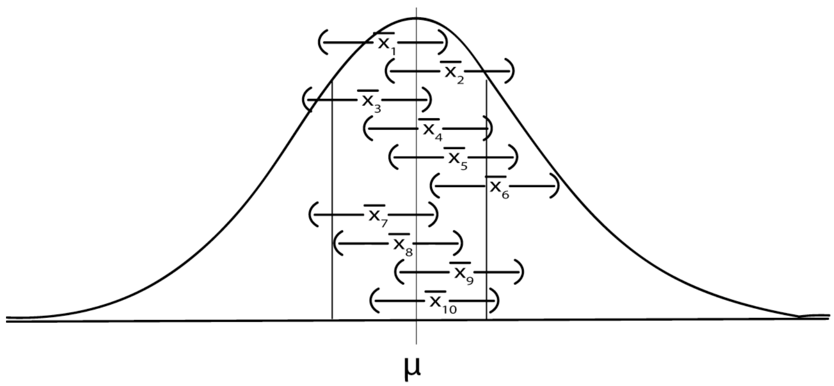
The interval image of confidence is the one often shown in introductory statistics texts. This image entails: (1) collecting a sample statistic, (2) finding the confidence interval around that sample statistic by adding and subtracting the margin of error from the sample statistic, and (3) repeating this process many times. In the interval image, the level of statistical confidence is defined as the percentage of intervals that capture the population parameter (see Figure 2).

Figure 2. Interval Image of Confidence.



The authors of this paper argue that the integration and coordination of both the interval image and the distribution image is essential for a coherent understanding of level of confidence. Figure 3 shows a representation of the *integrated image*. The integrated image places the unknown population parameter as a fixed, constant value. Through repeated sampling, sample statistics begin to clump around the population parameter, and the relationship between the percentage of sample statistics within a given distance of the population parameter *and* the percentage of confidence intervals that capture the population parameter is apparent. With this imagery, the rare event approach to statistical inference is more transparent. For example, a sample statistic not within a predetermined distance of the true mean and the confidence interval around that sample statistic does not capture the true mean, for instance, the sixth sample mean in Figure 3. Thus, this sample statistic is considered unusual in the sense that it is probabilistically less likely. Statistical inference is founded on the idea that obtaining a sample statistic that is unusual is a strong indication that the true population parameter may in fact be significantly different than the estimate.

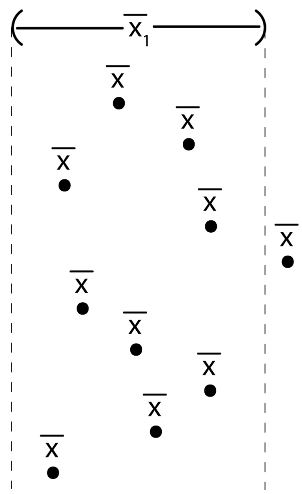
Figure 3. Integrated Image.



The "misconception" image of level of confidence

Using sampling distributions as a vehicle for constructing an understanding of statistical inference is not trivial. Students and high school teachers may experience difficulty: (1) reasoning about properties of sampling distributions such as shape and variability; and (2) reasoning about a distribution of sample statistics as distinct from a distribution of a single sample (Chance, delMas & Garfield, 2004; Heid et al., 2005; Saldanha & Thompson, 2007). There appear to be few studies investigating how, if at all, students relate sampling distributions to statistical confidence or hypothesis testing. The relatively little research that has been done in this area reveals a common misconception whereby the level of confidence is defined by the percentage of sample statistics that fall within the confidence interval constructed from the statistic computed from the first sample pulled (Cumming, Williams, & Fidler, 2004; Liu & Thompson, 2005). A visual representation of this "misconception" image is shown in Figure 4.

Figure 4. The “misconception”.



Despite the inherent difficulties in forming a coherent understanding of the role that sampling distributions play in defining statistical confidence, the statistics education community advocates that developing students’ knowledge of sampling distributions is a promising instructional strategy for assisting the development of informal inference skills. For example, the National Council of Teachers of Mathematics (NCTM, 2000) Data Analysis and Probability Standards for grades 9-12 recommend that teachers of statistics “use simulations to explore the variability of sample statistics from a known population and to construct sampling distributions” and “use sampling distributions as the basis for informal inference” (p. 325). The Guidelines for Assessment and Instruction in Statistics Education (GAISE, 2007) provide a similar recommendation; they argue that students should understand “the concept of a sampling distribution and how it applies to making statistical inferences based on samples of data” (p.7). An underlying premise throughout much the literature is that if students have the opportunity to construct sampling distributions from a known population and focus on the underlying concepts and relationships between the properties of sampling distributions and statistical inference, then students should be able to transfer this knowledge to other contexts where the population parameters are unknown, such as public opinion polls. The recommendations by these organizations and the arguments of statistics educators for a didactic model which uses sampling distributions as a basis for supporting students' development of informal inference skills is promising, yet the lack of empirical studies investigating these recommendations needs to be addressed. This report provides an initial investigation into the ways in which students apply their knowledge of repeated sampling and the construction of sampling distributions to polling contexts. In particular, this preliminary report presents two mini-case studies, focusing on the thinking of two undergraduate pre-service high school teachers enrolled in a statistics for teachers course.

DATA COLLECTION METHODS

This preliminary report focuses on two students, Kenny and Sara (pseudonyms) enrolled in a statistics for teachers course. Kenny and Sara were upper division undergraduate mathematics majors who were planning to become high school mathematics teachers. The statistics for teachers course is designed to stress important statistical concepts such as sampling distributions and their application to ideas of statistical inference. The course focused heavily on both physical hands-on simulations and computer simulations for building sampling distributions as a basis for informal inference. Thus, the students in this class had substantial classroom experiences working with sampling distributions from a known population. This report focuses on the ways in which Kenny and Sara coordinated their knowledge of repeated sampling and statistical confidence in relation to a polling scenario.

Data collection methods consisted of pre and post-surveys and pre and post-interviews. The surveys contained several polling scenarios where students were asked to provide their interpretation of polling results and their interpretations margin of error and statistical confidence in the context of the poll. Hour long clinical interviews were conducted in order to follow up on students' survey responses and provided the researchers with substantial depth into students' thinking. The first two authors independently examined the surveys and constructed a series of follow-up questions for the interviews. In addition, time was given for spontaneous follow-up questions during the interviews based on the responses of the participant. The interviews were videotaped and transcribed. Each author reviewed the transcripts and created summaries and visual representations of Kenny and Sara's thinking. The video was reviewed several times and the research team engaged in continued discussion of Kenny and Sara's thinking until consensus was reached. Three of the polling tasks from the post survey and the follow-up interview form the basis of the current report (see Figure 5).

Figure 5. Sample survey tasks.

(1) According to the latest New York Times/CBS News poll, Mr. Obama's approval rating is 63 percent. The national telephone poll was conducted the week of February 15 with 1,112 adults, and has a margin of error of plus or minus 3 percentage points.

Please explain what margin of error means to you in this context.

(2) According to a Gallup Poll conducted the same week as the New York Times/CBS poll, President Obama's approval rating is 67%. Results are based on telephone interviews with 1,551 national adults, aged 18 and older. They report 95% confidence that the maximum margin of error is ±3 percentage points.

What do you think the 95% confidence means in this context?

(3) Does it surprise you that the NY Times/CBS poll found a 63% approval rating while the Gallup Poll found a 67% approval rating? Explain why or why not.

RESULTS

Kenny and Sara both reasoned via the "misconception". That is, they both indicated that if the sampling process was repeated, then the level of confidence could be defined as the percentage of sample statistics that fall within the confidence interval created from the original sample statistic. Despite the fact that Kenny and Sara did not demonstrate a completely coherent understanding of statistical confidence, their reasoning was rather sophisticated and they were focused on sense making in the polling context. This section of the paper provides a deeper look into the "misconception", and how this view is mathematically reasonable albeit non-normative. There are three key points in the discussion of Kenny's and Sara's reasoning: (1) both assumed the first sample was representative and that the population parameter is never known; (2) both showed evidence of thinking from an interval image and a distribution image perspective, but neither had completely coordinated the two perspectives; and, (3) they differed in their opinions regarding what sample estimates would surprise them. Following the discussion of Kenny's and Sara's reasoning on these three points is an analysis of their discussion of what result would surprise them mediated through the lens of the "misconception" image and compared to a completely coherent image.

The sample will be representative and the population parameter in never known

Both Kenny and Sara repeatedly assume the first sample is representative and, subsequently, that the confidence interval created around the first sample statistic will capture the population parameter. This is evidenced in Kenny's discussion below:

> Kenny: What each poll is reporting is a sample statistic and it's supposed to represent the population parameter…On, on the whole margin of error idea, well our sample is a representative sample and we calculate our margin of error around this sample statistic that we get. …and we imagine that the population parameter is inside that interval

In the following excerpt, Sara further discusses how unlikely it would be for the first sample to produce a sample statistic that is not representative.

> Sara: Cause this is just an estimate. This is just a sample. A poll. A..you know it's a sample of the whole. …And you know to just randomly first time get one outside (*referring to the 5 or 10% outside the confidence interval*), that's like y-you know that's just. It's a little odd. You know you shouldn't get it the first time. It should, you know, 70 tries, and oh, my gosh, I got one or something.

The assumption that the sample and its corresponding statistic are representative is a reasonable assumption. In introductory statistics, students learn that the sample mean is a good estimate of the population mean, assuming the sample is representative. In addition, from a probabilistic perspective it is likely to get a sample mean that is within two standard errors of the population mean.

Kenny and Sara also recognized that the population parameter is not something that can ever be known. The two excerpts below show instances where they discussed how the population parameter is an unknown quantity.

> Kenny: And in, this population parameter is not something will ever really know.
>
> Sara: Well that's an estimate. They don't know if that's true or not. They're saying that the real one, or you know what we hope is the real one should be you know 60 to 66.

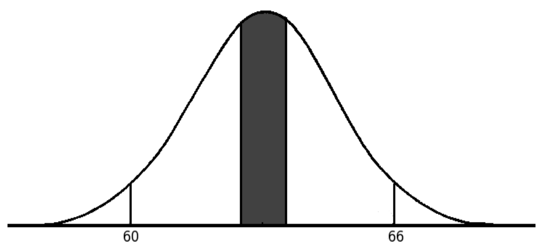
Kenny's and Sara's concept images of level of confidence

Kenny and Sara both articulated the "misconception" image of the level of confidence at the beginning of their interviews, namely as the percentage of sample statistics that fall inside the original confidence interval. Regardless of being non-normative, this image indicates that they both considered statistical confidence from a repeated sampling perspective. In addition, throughout the interview, both Kenny and Sara expressed images of the level of confidence focusing on the distribution of sample statistics. For instance, when Kenny initially constructed a sampling distribution during the post interview he placed the reported sample statistic in the center rather than the fixed unknown population parameter.

> Kenny: …so if you were to repeat this poll several hundred times you'd presumably get a number of different results. This number (*63% - the reported sample statistic*) would be, you know, you'd get a new one for each time you did the poll and we're assuming it would be normally distributed. …And the 63 would be somewhere in the middle here (*points to the shade region in Figure 6*). And we're using it because this is the one that they did.

In this case, we view Kenny's distribution as *theoretical* insofar as Kenny describes the result in terms of an expected outcome.

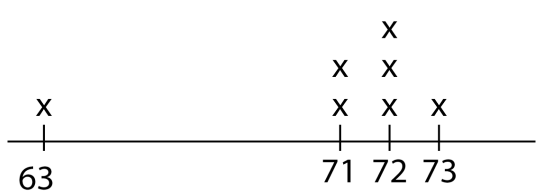
Figure 6. Kenny's (theoretical) distribution.



Given that Kenny recognized that the population parameter was a value that he would never know, the interviewers followed up by asking him how one might test the accuracy of a poll. The excerpt below provides evidence that Kenny imagined creating an *empirical* sampling distribution and looking for where the majority of the sample statistics clump together. If many of the sample statistics are clumped far away from the original sample statistic then one could reasonably assume that the original statistic is not representative.

> Kenny: If you were to repeat the poll then a few times, umm and see where your results come out. ...Let's say umm, someone does a poll and they come up with a 63% result and then someone else does a series of similar polls. They do them say 6 times or so and the results are really, they're like 71, 72, 72, 73, 71, 72% well that's enough of, of a difference where you could say maybe that first poll wasn't accurate for whatever reason.

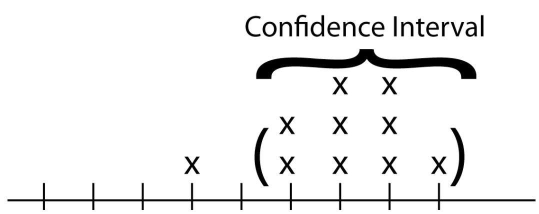
Figure 7. Kenny's (empirical) distribution.



Sara similarly discussed the formation of a confidence interval from both theoretical and empirical perspective. In the excerpt below, Sara first discusses how a *theoretical* confidence interval can be formed from a formula. The discussion (notably the second italicized sentence) also highlights Sara's "misconception" image of the level of confidence. The last italicized sentence in the excerpt indicates the formation of an *empirical* confidence interval, where she actually collects 10 sample statistics and determines the confidence interval by observing where the sample statistics are clumped within the empirical distribution (see Figure 8).

> Sara: You can have...*you can start with this 67 percent and you can form an interval around it using the errors and the formulas*…Anyways, you can form this ...you know this interval and you would expect if you took, you know, if it's 90 percent...*if you took 10, that you would expect 9 of them to be in here* 1 being out here floating. …*Or you know you can do 10 and find, you know, plot em and you could say oh 9 of them are here.* …It's just saying that, you know, in this case it would encompass, you know, 90 percent of these means or of the means of all the other samples. Put'em together.

Figure 8. Sara's empirical distribution.



The interviewer followed-up by asking Sara whether her two ways of forming a confidence interval would be the same and she seemed to indicate an expectation that the theoretical interval is larger.

> Interviewer: But are these intervals the same?
>
> Sara: Well it depends on what samples you have. Cause ...*this one* (points to the theoretical) *is probably gonna be larger cause it says, you know, it's done with a formula*. It says that …I can get 90 percent of the data in here or you know, 90 percent of the sampling means of a 100 should come here. *This* (*empirical distribution*) *says well I did...I did a, you know, 100 of them*. Well, you know, I got a 100 that are slightly skewed to the right, you know.

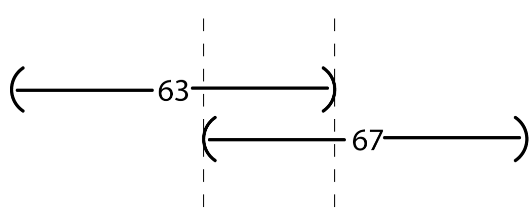
When asked about why she expects the theoretical one to be larger she says “Cause that takes into account all the possibilities”. This is an indication that perhaps Sara views the situation probabilistically and sees the confidence interval formed by the formula as accounting for all possible sample statistics.

What constitutes a surprising result?

Despite the fact that Kenny and Sara both expressed similar ways of thinking about statistical confidence, they had different perspectives on what values a new sample statistic could take on before it would constitute a surprising result in comparison to the original sample statistic. In the third sampling task (Figure 5), Kenny did not find the difference between the two approval rating polling results, 63% and 67%, surprising. His response, in the excerpt below, indicates that even though the approval rating of 67% does not fall in the confidence interval around the 63% (or vice versa), the 67% is not that far away from the interval around the 63% approval rating (see Figure 9).

> Kenny: I guess I’d be like amused, but not surprised. …Because its not that far outside the range. …You’d expect a few of those points to fall outside of that interval. You don’t expect them all to be within that interval.

Figure 9. Is it surprising that the 67% does not fall within the original confidence interval?



Sara, however, does find the difference between the 63 and 67% approval rating surprising for exactly the reason the “misconception” image would suggest. That is, if 95% of the sample statistics should fall in the interval around the first sample statistic (63% in this case), then only 5% fall outside that interval so the 67% is pretty unusual.

> Sara: But this 63 is not included in this 95 percent accurate interval. So it's the 10 or 5 percent that's outside and that just seems a little off. And then you know, this 67. It's not included in this other 90…95 percent confident interval.

On the one hand, Sara shows consistency in her image of the percentage of sample statistics that should be captured within the original confidence interval and her sense that if only 5% fall outside that interval then the 67% is in fact a bit surprising. On the other hand, although Kenny does not seem to be consistent between his view of statistical confidence and his notion of a surprising result, he does appear to have some flexibility in thinking of the original sample statistic as an estimate of the population parameter and thus some flexibility with his expectation of sample statistics with overlapping confidence intervals.

An analysis of the thinking of Kenny and Sara via the "misconception" image

Tables 1 and 2 provide a summary of the possible cases for the original confidence interval (OCI) and a new confidence interval (NCI) created during the hypothetical repeated sampling process. In Table 1, the sample statistic in the NCI falls within the range of the OCI. According to the "misconception" image this new statistic would be one of the 95 and should not be considered unusual. Both Kenny and Sara suggested that they would not be surprised by such a result, likely because they assume that the population parameter is inside the OCI. However, as Table 1 indicates, there are three ways in which such a view is incoherent: (1) the OCI does not capture the population parameter, but the NCI does, (2) both the OCI and NCI do not capture the population parameter, and (3) the OCI captures the population parameter, but the NCI does not. Thus, there are times when Kenny and Sara should consider some of the sample statistics unusual, but they do not.

Table 1. Four possible cases when the new sample statistic falls inside the OCI

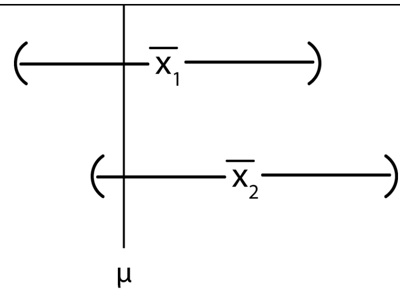
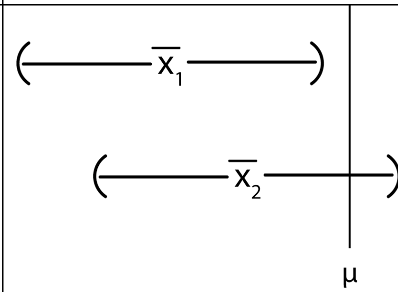
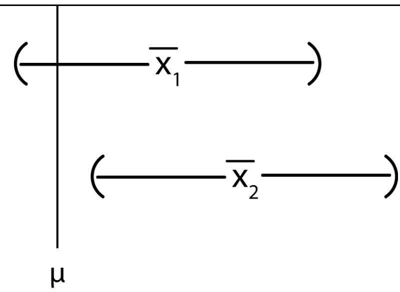
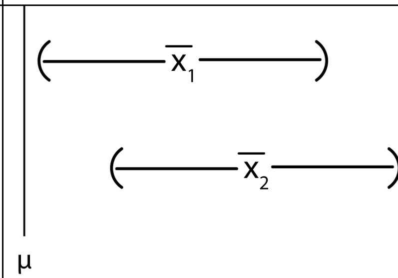
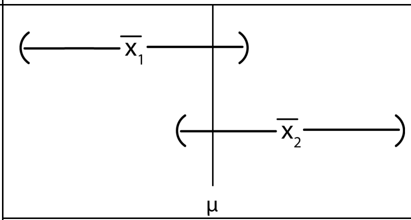
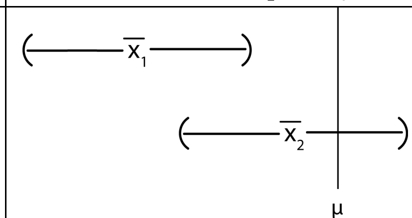
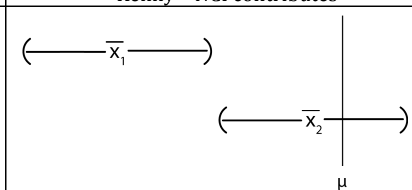
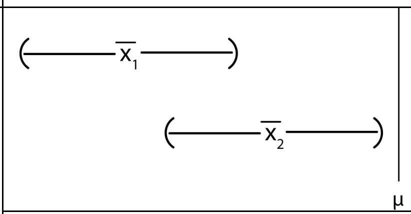
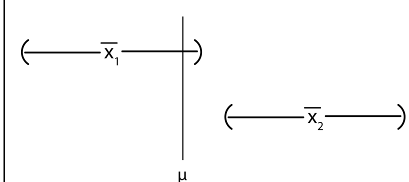
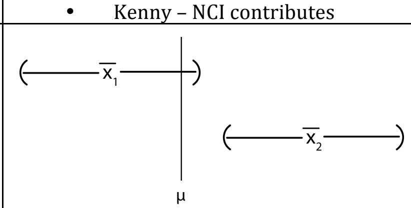
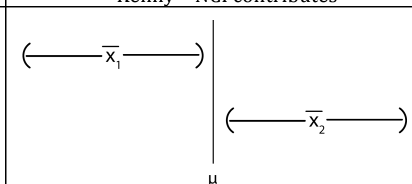
New Sample Statistic Inside Original Confidence Interval		
	Original Confidence Interval (OCI) Contains μ	**Original Confidence Interval (OCI) Does Not Contain μ**
New Confidence Interval (NCI) Contains μ		
	• Normative (OCI and NCI contribute to the level of confidence)	• Non-normative (OCI does not contribute to the level of confidence)
New Confidence Interval (NCI) Does Not Contain μ		
	• Non-normative (NCI does not contribute to the level of confidence)	• Non-normative (OCI and NCI do not contribute to the level of confidence)

Table 2, shows the possible cases when the new sample statistic does not fall within the OCI. According to the "misconception" image in each of these cases the new sample statistic should be considered unusual because it is not one of the 95 within the OCI. Sara seems to be consistent with this image when she discussed results that would surprise her. However, Kenny reasoned about surprising results in a slightly different way. He was not surprised as long as the

OCI and NCI overlapped. Yet, the only time when this image is coherent is when both the OCI and the NCI capture the population parameter.

Table 2. Seven cases when the new sample statistic falls outside the OCI.

New Sample Statistic Outside Original Confidence Interval			
		OCI Captures μ	**OCI Does Not Capture μ**
NCI Captures μ	**Two CI's Overlap**		
		• Normative • Kenny – NCI contributes	• Non-Normative • Kenny – NCI contributes
	Two CI's Do Not Overlap		
		• Can not occur	• Non-normative • Kenny – NCI does not contribute
NCI Does Not Capture μ	**Two CI's Overlap**		
		• Non-normative • Kenny – NCI contributes	• Non-normative • Kenny – NCI contributes
	Two CI's Do Not Overlap		
		• Non-normative • Kenny – NCI does not contribute	• Non-normative • Kenny – NCI does not contribute

Both Kenny and Sara will consider some sample statistics unusual when they shouldn't and others as not surprising when they are in fact unusual. Kenny's image of surprising is much less restrictive than Sara's. Kenny includes more than the 95 sample statistics that he believes should fall within the OCI into his image. In general, the further away the original sample statistic is from the population parameter, the more likely Kenny and Sara are to include unusual sample statistics in their image and the less likely they are to include sample statistics that should be included.

Sara and Kenny do appear to have an interval image and a distribution image, but have not fully integrated the two pictures. They place the original sample statistic in the center of the distribution. If the original statistic is “close” (for example within one standard error) of the population parameter then their images are not terribly problematic. Since the distribution centered around the original sample statistic captures most of the sample statistics that are within 2 standard errors of the population parameter. However, the farther the original sample statistic is from the population parameter then the more problematic Kenny’s and Sara’s images become because the confidence interval around the original sample statistic will capture fewer of the sample statistics within 2 standard errors of the true population parameter. Their images seem to ignore that the population parameter is a fixed, unknown constant that should be placed at the center of their distribution image.

CONCLUSIONS

It appears inherently problematic to coordinate the distribution image and the interval image, placing the unknown population parameter in the center of the distribution to form a coherent conception of confidence interval. In part it seems that the unknown population parameter is abstract and elusive and therefore students may find it difficult to think about its placement in their image of statistical confidence. It is intellectually an easier task to place the first sample statistic in the center of the distribution and to think of it as a good representative of the population parameter. This seemingly straightforward and largely reasonable substitution makes the coordination of the distribution and interval images inherently difficult.

Our results suggest that the ability for students to transfer their knowledge of sampling distributions and ideas of informal statistical inference from a known population (e.g., a jar of white and black beans) to other contexts, such as public opinion polls, is neither immediate nor trivial. This result seems particularly significant in light of the fact that these students had substantial mathematics backgrounds. How do we help future teachers see the conceptual differences between the misconception view and a completely normative view? Especially when the "misconception" view is more natural and makes a whole lot more sense. The need for future research investigating didactic methods for supporting students in coordinating ideas of sampling distributions to polling contexts is tremendous, especially if statistics educators are to help students develop robust ideas of inference in contexts they are likely to come across on a daily basis.

References

Aliaga, M., Cobb, G., Cuff, C., Garfield, J., Gould, R., Lock, R., Moore, T., Rossman, A., Stephenson, B., Utts, J., Velleman, P., & Witmer, J. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) college report*. Alexandria, VA: American Statistical Association. (Also available at http://www.amstat.org/)

Chance, B., delMas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In J. Garfield & D. Ben-Zvi (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 295-324). Dordrecht, The Netherlands: Kluwer Academic.

Cumming, G., Williams, J., & Fidler, F. (2004). Replication, and researchers' understanding of confidence intervals and standard error bars. *Understanding Statistics, 3*, 299-311.

Gal, I. (2003). Teaching for statistical literacy and services of statistics agencies. *The American Statistician 57(2)*

Garfield, J.B., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Springer.

Garfield, J.B., delMas, R., & Chance, B. (2007). Using students' informal notions of variability to develop an understanding of formal measures of variability. In M. Lovett and P. Shah (Eds.), Thinking with Data (Proceedings of the 33rd Carnegie Symposium on Cognition). (pp.117- 147). New York: Erlbaum.

Heid, K. M., Perkinson, D., Peters, S.A., & Fratto, C.L. (2005) Making and managing distinctions – the case of sampling distributions. *Proceedings of the 27th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education.*

Konold, C. (1989). Informal conceptions of probability. *Cognition and Instructions, 6*, 59-98.

Liu, Y. & Thompson, P. (2005). Understandings of margin of error. In Lloyd, G. M., Wilson, M., Wilkins, J. L. M., & Behm, S. L. (Eds.). *Proceedings of the 27th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education.*

Luzter, D.J., Rodi, S.B., Kirkman, & Maxwell, J.W. (2005). *Statistical abstract of undergraduate programs in mathematical sciences in the United States: Fall 2000 CBMS Survey*. Retrieved on 5/17/05 at http://www.ams.org/cbms2000.html

National Council of Teachers of Mathematics. (2000). *Principals and standards for school mathematics*. National Council of Teachers of Mathematics, Inc., Reston, VA.

Saldanha, L., & Thompson, P. (2007). Exploring connections between sampling distributions and statistical inference: An analysis of students' engagement and thinking in the context of instruction involving repeated sampling. *International Electronic Journal of Mathematics Education, 2(3),* pp. 271-297.

Tall, D., & Vinner, S. (1981). Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational Studies in Mathematics*, 12, pp. 151–169.

Watson, J. M. and J. B. Moritz (2000). The longitudinal development of understanding of average. *Mathematical Thinking & Learning 2(1&2)*, pp. 11-50.