

Informal Inferential Reasoning: Interval Estimates of Parameters

Theodosia Prodromou¹

¹ School of Education, University of New England, Australia

Correspondence: Theodosia Prodromou, School of Education, University of New England, Australia. Tel: 61-2-6773-3237. E-mail: theodosia.prodromou@une.edu.au

Received: March 26, 2013 Accepted: April 16, 2013 Online Published: April 28, 2013

doi:10.5539/ijsp.v2n2p136 URL: <http://dx.doi.org/10.5539/ijsp.v2n2p136>

Abstract

This research examined the informal inferential reasoning of senior secondary school students (age 17) when engaged in a computer-simulated sampling activity calling for the estimation of population parameters. The students undertook a task involving interval estimation of parameters within a computer-simulated environment. The research observed the students while they made and then explained their parameter estimates in order to better understand how the students formed the interval estimates, with particular attention to different strategies they adopted in forming these estimates. Activities involved sampling and estimating across three different sample size situations followed by a reflection stage to compare the estimates. Results of the analysis of the discussion between the students and the researcher during the students' activities are presented. A number of strategies for forming an interval estimate emerged. The students experimented with choosing different strategies for forming the interval estimate when a new sample (observed values) was drawn. The research findings are useful for informing the teaching of interval estimation to school-aged students.

Keywords: informal inferential reasoning, interval estimation, parameter estimates, population, samples, sampling

1. Introduction

Informal Inferential Reasoning is the process of drawing generalised conclusions from data. Four key principles have been identified as important to making informal inferences from data: "(1) Generalization, including predictions, parameter estimates, and conclusions, that extend beyond describing the given data; (2) the use of data as evidence for those generalizations; (3) employment of probabilistic language in describing the generalization, including informal reference to levels of certainty about the conclusions drawn." (Makar & Rubin, 2009, p. 85); and (4) comparison of datasets with a model (Bakker, Kent, Derry, Noss, & Hoyles, 2008).

Making inferences informally gives students a sense of the power of statistical techniques for making reasoned judgments and decisions about data drawn from real-world contexts.

For the first critical principle of making informal inferences, generalising beyond data, much of the research into students' reasoning has focused on drawing conclusions (e.g., Bakker & Derry, 2011; Pfannkuch, 2005; Watson & Moritz, 2000a) and making predictions (e.g., Makar & Rubin, 2009; Prodromou, 2011). The estimating parameters competence has not been as well researched to date and this has prompted the research shared in this paper.

This paper reports exploratory research into how students reason when estimating population parameters, in particular when making interval estimates of the parameters. The strategies used by students to create interval estimates and the developing conceptions of the process of sampling were of interest to the researcher.

2. Estimating Parameters by an Interval Estimate

When sampling and generalising beyond the data, statistics may be used to estimate unknown population parameters. Estimation of parameters is associated with the process by which one makes inferences about a population based on information gained from a sample.

Parameters can be estimated by providing a point estimate or an interval estimate. One can use sample data to make an "interval estimate" that is to calculate an interval of possible (or probable) values for an unknown population parameter within which the actual value of the population parameter lies.

Formally, the interval estimate is expressed as a sample statistic plus or minus a margin of error, which provides the range of values around the sample statistic that "may" encompass the true value of the population parameter.

So, interval estimates provide a margin of error but not the confidence level.

It is of paramount importance to introduce students to *estimating with confidence* but there has been little research to date about this. One important proposal for such teaching has been provided by Rossman (2008, p. 18) who proposed examples for teaching that involved categorical variables, because “such variables provide a simpler context in which students can focus on key ideas of inference.” He suggested a simulation (about kissing couples and whether those couples lean to the right or left) of randomization tests as an informal way to “introduce students to the logic of statistical inference” (p. 17). The simulation allowed students to investigate the plausibility of values for the parameter proportion (of couples leaning to the right) other than 0.5 by repeating the simulation analysis with their chosen values. Rossman’s (2008, p. 16) strategy was to have students “reject any value of the population proportion that puts the observed data in the tail of its sampling distribution”. He suggested that having students compare different models (changing the population proportion) via simulation (of 1000 repetitions each) can increase students’ informal reasoning abilities about interval estimation and estimating with confidence, and introduce students to the key role played by chance variation in statistical inference. Crucial to this reasoning is the notion that parameter estimates of an unknown population change every time a new iteration of a simulation is performed in spite of the fact that each simulation is run under the same conditions. In fact, this becomes the central catalyst for inferential methods.

Statistics educators have enabled students to experience sampling variation by conducting a study and calculating an estimate (see, e.g., Chance, delMas, & Garfield, 2004) while appreciating that estimates and other statistics (e.g., sample means, proportions) change every time a new study is performed, even if each study is performed under exactly the same conditions. Tasks which involved study-to-study variation in estimates provide an effective way to introduce students to the logic of statistical inferential methods that incorporate uncertainty, and emphasize the key role played by chance variation in statistical inference that involves uncertainty as the significance test and confidence interval (Wild, 2006).

3. Understanding Sampling and Related Concepts

Sampling, critical to prediction and decision making in many aspects of life, is fundamentally connected to the notion of making inferences about populations (Watson, 2006) and being able to reason while sampling depends on an understanding of sampling and related concepts. *Sampling* is the “act, process, or technique of selecting an appropriate sample” (Farlex, 2012), or, more specifically, selecting a representative part of a population for the purpose of determining parameters or characteristics of the whole population. Two concepts central to the understanding of sampling are sample and variation. *Sample* is defined as a “portion, piece or segment that is representative of a whole” (Farlex, 2012). The statistical concept of sampling, or taking a sample, is different from the colloquial notion of taking a sample, such as when tasting food, and the most important concept in helping students to understand the difference between the statistical concept of sample and the colloquial is the variation (Watson, 2006). *Variation* is the “act, process, condition or result of changing or varying” (Farlex, 2012) and the main purpose of the sample is to represent the variation in a heterogeneous population. Understanding the sampling concept also depends on understanding other statistical concepts, including distribution, randomness, likelihood, and representativeness (Ben-Zvi, Aridor, Makar, & Bakker, 2012; Watson & Moritz, 2000b). For a more detailed understanding of sampling, an understanding of the various sampling techniques that can be used would also need to be developed (Watson, 2004).

Although literature is replete with past research on tertiary students’ conceptions of sampling, limited research has been undertaken until recently on school students’ conceptions of sampling (Garfield & Ben-Zvi, 2008). Now we will move on to discuss relevant research on students’ conception of sampling, samples, and variation.

Early research into school students’ conceptions of sampling focused on two central ideas. One central idea was the *potential bias from drawing conclusions about populations based on small samples*. Young students (aged 8 to 9) have been shown to have relatively naïve conceptions about samples (Watson & Moritz, 2000a, 2000b). These students were typically comfortable drawing conclusions about a population based on small samples without recognizing any potential problems of bias. Even early middle school students (aged 13 to 14) understood the concept of sample in real world situations, but had difficulties making the transition to the formal statistical meaning and using appropriate associated terminology (Watson & Moritz, 2000b). Admittedly, given that adult students who take statistics courses often have trouble understanding formal statistics and terminology, the difficulty of this transition is not entirely surprising.

The second central idea was the *difference between drawing small samples from a homogenous population and*

from a heterogeneous population to make conclusions about the population from which the sample was drawn. Older students (age 14 to 15), who did show concern for potential errors arising from small samples, were not able to generalize ideas inherent in sampling from homogenous entities (e.g., a small sample of blood) to the notion of sampling variation and the need for large samples when making inferences from data (Watson & Moritz, 2000b).

More recently, research has also focused on a third central idea, *distinguishing between sample and population* (Pfannkuch, 2008), which is critical to the two central ideas described above. Research into developing a better understanding, from a pedagogical perspective, of how to approach differentiating between sample and population provides useful insights into student understanding of core concepts. One such approach, Growing Samples, originally suggested by Konold and Pollatsek (2002), was formally developed by Bakker (2004), and is now widely used (see, e.g., Ben-Zvi et al., 2012; Prodromou, 2011). This approach, based on the integration of core statistical concepts including sampling and variation, was developed to help students reason about sampling in a context of variability (Bakker, 2004). As part of this approach Bakker helped students (aged 13 to 14) engage with a sequence of “growing samples” activities to see stable patterns generated by larger samples, thus better understanding that larger samples are less variable and better represent a population.

Research by Ben-Zvi et al. (2012) focused on a similar approach, using *TinkerPlots*, with students (aged 11) drawing conclusions about the population first from a sample of size eight from their class (including themselves), moving to a bigger sample (a whole class) and then to the whole grade in the school. The students experienced the limitations of small samples when making inferences about a larger population, including making emerging quantification of confidence for such inferences and building interconnections with the key concept of variation while reasoning about sampling. Prodromou (2011) worked with older students (aged 14 to 15) who were engaged in prediction and noticed that these students made interval predictions rather than providing a single value, which can be taken as an indication that they lacked the confidence to make a single point estimate. Importantly, these students generally recognized the relationship between sample size and the confidence interval for a given confidence level.

This research picks up on the concerns of the scholars mentioned (e.g., Bakker, 2004; Ben-Zvi et al., 2012; Pfannkuch, 2008) about how students understand sampling, and how to teach sampling, by focusing on the way school students attempt to develop an understanding of the process of estimating population parameters.

4. Aim

This research investigated how senior secondary school students reason when making informal inferences from data. The students engaged in a statistical inference task that involved forming interval estimates of the percentage of a population species, that is a population parameter, within a computer-simulated environment. Following is a description of the interval estimation task, the cohort of participants, and the method used.

5. Interval Estimation Task

The interval estimation task was based on information provided by use of a computer-based simulation titled “Murphy’s Dam”. The simulation, presented in a spreadsheet, introduced a scenario in which the owner of a dam stocked with three fish species (Bass, Perch, & Trout, Figure 1) wanted advice about the percentage of perch in his dam. The spreadsheet simulated drawing a sample of fish from the dam and displayed both the number and percentage of each type of fish in the sample (e.g., see Figure 2). The spreadsheet contained three separate sheets, one each for drawing samples of size 20, 50 and 100. The sampling was described as drawing with replacement. The task had four stages: Sample Size 20, Sample Size 50, Sample Size 100, and Reflection.

Murphy’s Dam

Brian Murphy has a dam, on his farm, which contains many fish of three different species: Bass, Perch and Trout. Since introducing each of the three species the number of fish has grown considerably. Brian would like to estimate the percentage of perch he now has in the dam.

You have been consulted to provide this advice to Brian. Your estimate of the percentage of perch will be based on a sample of fish that you draw (catch and release) from the dam.

Your sampling of fish will be based on the assumption that you are drawing the fish in such a way that each fish (no matter what species) is equally likely to be caught.

Figure 1. The Murphy’s Dam Scenario, as presented to students

Murphy's Dam			
Draw a sample of 50 fish			
Species	Bass	Perch	Trout
Number	13	27	10
Percentage (%)	26	54	20

Figure 2. Display showing the number and percentage for the three species of fish for a single sample

In the *Sample Size 20 Stage* students engaged in estimating the percentage of perch in the dam by providing an upper and lower value (i.e., an interval) for the percentage after drawing samples of 20 fish. Students were allowed to draw ten separate samples from the dam. After each draw, the students made an interval estimate and were asked to explain why they gave those particular limits for the interval estimate. They were provided with a recording sheet (Figure 3) to record the observed percentage for each species of fish caught and to record the interval estimate of the percentage of perch as a lower limit and an upper limit. Students verbally justified their interval estimates.

Fish caught from the dam			Interval estimate of percentage (%) of perch	
Bass (%)	Perch (%)	Trout (%)	lower limit	upper limit

Figure 3. Recording sheet (partial; figure shows two of the ten recording lines)

The *Sample Size 50 Stage* and the *Sample Size 100 Stage* involved the same steps as the *Sample Size 20 Stage* except that larger samples (of size 50 and 100, respectively) were drawn. The *Reflection Stage* involved the students reasoning about what had occurred in the first three stages by comparing the interval estimates of the percentage of perch.

6. Methodology

6.1 Participants

The statistical inference task was undertaken by six average-ability male Year 11 (age 17) students studying Mathematics General in an Australian secondary school. Participation was voluntary and students chose their own partner for the task. Final choice of which pairs participated was made by the teacher so as to include those most likely to be able to articulate their reasoning. The students had been taught sampling previously, so they were experiencing tasks that should have been familiar to them. The pairs of students undertook the activities out of class time.

The researcher was a participant observer when students were completing the task. The role of the participant observer was to interact with the students in order to probe the reasons or intuitions that might explain their actions.

6.2 Data Analysis

Data collected for the study included audio recordings of the students' voices as they worked, the worksheets completed by the students, and notes indicative of the students' expressions, gestures, and body language. The data analysis for this study focused on the verbal expressions captured in the audio recordings. The plain accounts (transcribed audio recordings) were analysed to infer explanations for students' actions and articulations. Progressive focusing analysis (Robson, 1993) was employed to determine strategies used to create the interval estimate and the reasoning about sampling that was involved when justifying their estimates. First, the recordings were transcribed and screenshots were included as necessary to make sense of the transcription. Then the author selected some of the sections that more clearly demonstrated the students' reasoning.

7. Results

Three pairs of boys undertook the task. Of these three pairs, Neil and Tim (pseudonyms) were the best at interacting with each other and articulating their thinking. Moreover, Neil and Tim articulated most of the ideas articulated by the other two pairs. Thus, the results are presented by way of the case of Neil and Tim, working through the four stages. For each stage, the most informative sections of the transcript are provided to illustrate the ways the students created their interval estimates and justified their estimates. In the transcripts N = Neil, T = Tim, and R =

Researcher. The samples and estimates are numbered according to their position in the sequence of the ten samples drawn, e.g., sample #5 was the fifth sample drawn and estimate #7 was the seventh estimate formed.

7.1 Sample Size 20 Stage

The recording sheets completed by Neil (Figure 4) and Tim (Figure 5) show the observed values and the estimated values for the ten samples drawn. Although Neil and Tim were asked to explain all ten interval estimates, only the reasoning for the estimates of most interest to the discussion is reported. For four of the samples drawn (#3, #4, #6 & #8) relevant parts of the working are described.

When Neil and Tim drew sample #3 they caught 35% Bass, 60% Perch and 5% Trout. Neil reasoned as follows:

1. N: I've gone for a lower limit of around 30% and an upper limit of around 72%... Because that correlates with my earlier percentages.
2. R: How does it correlate with the earlier percentages?
3. N: Because got 27 and 51... Then that would be, ah actually that would be, the lower limit would be about 31 probably. And that is in between 27 and 35. My upper limit is 72, which is between 51 and 88. And that fits together quite nicely.

Fish Caught from the dam			Interval estimate of percentage (%) of perch	
Bass(%)	Perch(%)	Trout(%)	lower limit	upper limit
30	40	30	27	51
20	70	10	35	88
35	60	5	31	72
25	65	10	32	77
25	50	25	29	65
50	30	20	19	46
55	20	25	13	39
35	45	20	28	58
45	35	20	24	48
15	55	30	30	69

Figure 4. Neil's recording sheet when drawing samples of size 20

Fish Caught from the dam			Interval estimate of percentage (%) of perch	
Bass(%)	Perch(%)	Trout(%)	lower limit	upper limit
30	40	30	17	50
20	70	10	35	88
35	60	5	30	80
25	65	10	32	75
25	50	25	28	67
50	30	20	19	45
55	20	25	11	39
35	45	20	27	58
45	35	20	22	47
15	55	30	27	73

Figure 5. Tim's recording sheet when drawing samples of size 20

Neil's explanation for estimate #3 (see line 3) was that he formed a lower limit estimate (30) between two previous lower limits, those for estimate #1 (27) and estimate #2 (35). He did not explain, however, why he chose to form his estimate by placing it between two previous estimates. He reasoned similarly when he chose an estimate for

the upper limit of the interval. Both Neil and Tim then began to look more closely at the percentage of fish caught and the interval estimates of percentage of Perch.

When Neil and Tim drew sample #4 they caught 25% Bass, 65% Perch, and 10% Trout. Neil reasoned as follows:

4. N: For the lower limit are 33.
5. R: Why 33?
6. N: Well, my last one was 31, and this is only 5% bigger... So it's not a huge jump, but it's noticeable to the 33 there.
7. R: Why didn't you go back to 34 or 32? I mean down from 35.
8. N: Because ...
9. R: Sorry, what was the previous one?
10. N: The last one was 31.
11. R: And you go to 33 ...
12. N: I've gone to 33.
13. R: Why didn't you go up by one ... and you go.
14. N: I had 32 and it wasn't happening so I changed it.
15. R: Why?
16. N: Don't know.
17. R: It's not too much difference.
18. N: It's probably a better amount. 32 is probably a better amount of difference.
19. R: Why?
20. N: Probably going up 5. When I went up 10 I had 5 difference. So that was each percent was 2 ... It's going up to 5... My upper limit is 77.
21. R: Why? And how did you go up? How did you change the previous #3 estimate? What was the previous #3 estimate?
22. N: My last one was 72... That's 5 % higher, and the one before that I had 88.
23. R: Yeah.
24. N: So I've just placed it roughly in between those two.

Neil focused on one previous estimate (#3) to make his new estimate #4. As part of his reasoning he tried to explain his decision about how much bigger to make the new estimate in relation to the previous estimate. Neil calculated that the difference between the #4 observed percentage (65) and the previously (#3) observed percentage (60) was 5 so he worked with the "5" increase. When Neil expresses the difference as "5% bigger" he is referring changing the value of the variable "percentage of perch in the dam" by 5. Here, and in many places through the transcript where someone refers to a percentage increase/decrease, the speaker is not referring to the percentage change in the variable, but to the absolute change in the variable. It is not clear, however, how Neil actually used this increase of "5" to decide on the increases needed to make his new estimates, although Neil did keep his sights on the 5% increase (lines 10, 20, and 22) when making both the lower and upper limits for the new #4 estimate.

For the lower limit of the #4 estimate Neil initially chose 32 but as he formulated his explanation he argued that it should be changed to 33 because "it wasn't happening" (line 14). It is not clear what "wasn't happening" meant, but Neil went on to explain how a change of 10 in the percentage represents "2" (line 20). It should be noted here that 10% of 20 (the sample size) is "2" but this is not how Neil explained it.

When Neil was asked by the researcher (line 21) to explain the upper limit of the #4 estimate (77), Neil referred back to the previous #3 estimate (72) and to another previous estimate #2 (88) and concluded that he has placed his new estimate "roughly in between these two" (i.e., estimate #3 and estimate #2). This last explanation suggests that he was reverting to a similar process that he used for estimate #3, which was placed between two previous estimates. It should be noted that, earlier on, even though the researcher mentioned the previous estimate #2 when probing for an explanation of the value (see line 7) selected for the new lower limit, Neil did not revert to considering the place of the new estimate between two previous estimates as he finally did for the upper limit.

When Neil and Tim drew sample #6 they caught 50% Bass, 30% Perch and 20% Trout. Their reasoning follows:

25. T: Well um, I put 19 as the lower limit.
26. R: Ok.
27. N: I got 19 as my lower limit.
28. R: Yes.

29. N: And my higher limit, my upper limit is 46. Because my, the 40% my upper limit was 51. So I've gone 46, it will be around 46, 46. I've gone 46.
30. T: Actually I've got 45.
31. R: In comparison to the previous estimates.
32. N: That one, I based it on.
33. R: You know here, did you have a look at the previous lower limit and upper limit?
34. N: Yeah. Upper limit.
35. R: Did you take into account an estimate for that one?
36. N: Yeah, I considered that one which was 40 and like 27 would have to be lower than that, probably 19. And my lower limit I'm making actually lower than normal, and my upper limit would be about 46.
37. R: Ok. Why do you have 45?
38. T: Well, I sorta took the upper limit of the first one into consideration, seeing that one's 50, and there is a 10% sorta difference so I took it down to 45 so it balances and makes more sense. And the lower limit, that I still base it off that one, except I know that was way too low so I just put it up by 2, cause that's the 80%, 40% should have been higher.

When forming estimate #6, Neil considered (line 36) the previously observed #1 percentage (40) and the related interval estimate #1 for Perch (lower limit 27 and upper limit 51) to make the new interval estimate with a lower limit of 19 and an upper limit of 46. His explanation was to have a new estimate that was lower than the previous estimate but the listener is left to assume that this is because the new observed value (30) is lower than observed #1 (40). A possible interpretation of Neil's statement that he is making the estimate "actually lower than normal" (line 36) is that he has taken more than just estimate #1 into consideration. However, there is insufficient explanation to determine whether he was using the word "normal" to refer to estimate #1 that he chose for his focus, or was actually looking at more than one of his previous lower limit estimates to get a feel for what was "normal". Tim similarly focused on the estimates associated with observed #1 percentage to form his estimate #6, and he provided a mathematical explanation for the amount of the new estimated limits again explaining that they change by "each percentage was 2" (line 38), as he did for estimate #3.

When Neil and Tim drew sample #8 they caught of 35% Bass, 45% Perch and 20% Trout. Their reasoning follows:

39. N: I have my lower limit. It is 28%...And my upper limit is 58%.
40. R: Ok. Why?
41. N: It's...45 is my lower limit for 40 was 27%. My lower limit for 50 was 29%, so I'm going in between those two. My upper limit for 40% was 51, and my upper limit for 50% was 65. I've just gone between those, tried to roughly guess between those two again, got 58.
42. T: Um, my lower limit was 27, so pretty much between 40 and 50%. And that's sort of the same with my upper limit which was 59.

When forming estimate #8 Neil considered two previously observed values (#1 and #5) which were either side of the current observed value (i.e., 40 and 50 are either side of "45") and then the estimate #8 was made between the estimates for these two previously observed values. So when Neil made the lower limit estimate #8 (28) he explained, pointing at his previous estimates of 27 and 29, "I'm going in between those two" (line 41), referring to 27 and 29. There is no evidence that this is a calculation of the average of 27 and 29. He simply stated that the number had to be between 27 and 29. Similarly, for the upper limit he chose between 51 (estimate #1) and 65 (estimate #5) and made his estimate "58" which is the average but again is not called that by Neil who explained "I've just gone between those" (line 41). When a new value is placed between just two previous values it is not possible to tell whether the student wants an average value unless this is specified.

7.2 Sample Size 50 Stage

For the second stage, Neil and Tim were again asked to explain all ten interval estimates but only the reasoning for the estimates of most interest are reported. For six of the samples drawn (#2, #3, #4, #6, #8 & #10) relevant parts of the working are described.

Fish Caught from the dam			Interval estimate of percentage (%) of perch	
Bass(%)	Perch(%)	Trout(%)	lower limit	upper limit
30	50	20	30	71
30	44	26	26	65
26	54	20	37	79
22	56	22	39	83
30	52	18	34	75
30	44	26	26	65
22	52	26	34	75
24	58	18	42	88
24	50	26	30	71
36	40	24	21	59

Figure 6. Neil's recording sheet when drawing samples of size 50

Fish Caught from the dam			Interval estimate of percentage (%) of perch	
Bass(%)	Perch(%)	Trout(%)	lower limit	upper limit
30	50	20	34	77
30	44	26	31	68
26	54	20	38	78
22	56	22	40	79
30	52	18	36	76
30	44	26	27	68
22	52	26	35	74
24	58	18	43	82
24	50	26	30	70
36	40	24	26	61

Figure 7. Tim's recording sheet when drawing samples of size 50

When Neil and Tim drew sample #2 they caught 30% Bass, 44% Perch and 26% Trout. Their reasoning follows:

43. N: For that one there, it was 26, was the lower limit and the upper limit was 65.

44. T: There up from mine, 31 and my upper was limit was 68.

45. R: And why guys? Why and did you select that limit?

46. N: I did it for the 50 one, my lower limit is 30. My upper limit is 71... Both numbers 20 outside. Keeping room for error... Just trying to base it off that.

47. T: Yeah, and my lower for the other one was 34 and the upper was 77 so I was just trying to base it off that as well.

When forming estimate #2 part of Neil's explanation was to revert to the process he used for estimate #1, which was to choose limits either side of the observed value. The reason was better explained here as choosing "both numbers 20 outside keeping room for error" (line 46). This is the first mention of any sort of balancing either side of an observed value.

Neil and Tim drew sample #3 (26% Bass, 54% Perch and 20% Trout). Their reasoning follows:

48. N: My lower limit was 37, and my upper limit was 79.

49. T: My lower limit was 38, and my upper was 78 as well. I was just trying to balance it and go off the first one as well, made sense.

50. N: Just made it, you know just, mine was 4% higher. So the upper limit was higher, the lower limit was up as well.

When forming estimate #3 Tim mentioned using two approaches. The first approach, similar to that taken when

making estimate #2, was to choose numbers either side of the observed value as the limits, as indicated by “trying to balance it” (line 49). The second approach, similar to that taken when making estimate #6 (Sample Size 20), was to choose a “suitable” previously observed value, in this case the first estimate made, and base the new interval estimate on the estimate related to the previous estimate, as indicated by “go off the first one” (line 49).

After drawing sample #4 (22% Bass, 56% Perch and 22% Trout) Neil and Tim reasoned as follows:

- 51. N: Lower limit of 39, and upper limit of 83.
- 52. R: Well, why did you choose those limits?
- 53. N: 56, 2% higher than 54. Then my lower limit comes up a bit, my upper limits come up a smudge as well.
- 54. T: Yeah my lower is 40 and my upper is 79 so it's only gone up a little bit too, because it is only 2% differential.
- 55. R: So you transfer the limit to percentages.
- 56. N: Yes. And we hope we didn't get it wrong the first time.

When forming estimate #4, Neil and Tim both made use of one previous estimate #3. The related observed value #3 (54) was conveniently chosen because it was numerically close to observed value #4 (56). Neil then decided to go up by “a smudge” (line 53), apparently not doing any mathematical calculation of how much to go up but assuming that if the increase in the observed value was small, then only a small change should be made in the interval limits. Tim took a similar approach. His expression “differential” (line 54) was only used this once and so may have been a corruption of the term difference rather than a special tool devised for dealing with the estimation process.

During their reasoning there is often a sense that Neil and Tim view their estimates as being either right or wrong. For example, Neil indicated that he was concerned about how correct the previous estimate was (line 56). One is left to assume that he is worried about how correct his new estimate was, if the previous estimate on which it was based was actually “wrong”.

When Neil and Tim drew sample #6 they caught 30% Bass, 44% Perch and 26% Trout. Neil's reasoning follows:

- 57. N: Oh, it's, we did 44 earlier, so some of the data has repeating samples.

Both Neil and Tim used the same estimate as for the observed value #2, which also had an observed value of 44% Perch. Neil used the term “repeating samples” (line 57) to indicate that the same percentage of Perch had been observed as earlier. It should be noted that the percentage of Bass and Trout were also the same as for observed value #2 which may have contributed to the notion of a repeating sample.

And after sample #8 they reasoned as follows:

- 58. N: My lower limit was 42 and upper limit was 88.
- 59. T: Lower are 43 and upper are 82.
- 60. R: Um did you use the same strategies as before?
- 61. T: Yeah.
- 62. N: Our numbers are getting very close.
- 63. R: Why do you think your numbers are getting very close?
- 64. N: Just getting very similar.
- 65. R: Why?
- 66. N: We are both smart guys, it must be the answer.

By the time Neil and Tim were forming estimate #8, they had provided almost identical estimates. Neil noticed that the estimates were similar (line 62) but when asked to explain why this might be, he said “just getting very similar” (line 64) but jokingly added “we are both smart guys, it must be the answer” (line 66).

After sample #10 they reasoned as follows:

- 67. N: Lower limit was 21, my upper limit was 59.
- 68. T: My lower limit was 26, my upper limit was 61.
- 69. R: Ok, like before?
- 70. N: Yeah, 40. The lowest starting number we had was 44 and that was 26 and 65 so I've just pulled it back a bit.
- 71. T: Yeah, I agree like mine, the last 44 we had I put as 31 and 68 so I just bring it down.

After noticing that the observed value for perch was lower than it had been for any of the previous samples, Tim

formed his estimate #10 by reducing a previous estimate “so I just bring it down” (line 71). There is no indication that the amount of the reduction has any mathematical basis. Neil used the term “starting number” (line 70) to refer to the observed percentage of Perch. This suggests that he viewed the observed number of Perch as the first number(s) he should consider when making an estimate because he actually referred to the entire column of sample percentage of Perch as the starting numbers.

7.3 Sample Size 100 Stage

Neil and Tim were again asked to explain all ten interval estimates but only the reasoning for the estimates of most interest are reported. For four of the samples drawn (#1, #2, #3, & #8) relevant parts of the working are described.

Fish Caught from the dam			Interval estimate of percentage (%) of perch	
Bass(%)	Perch(%)	Trout(%)	lower limit	upper limit
28	55	17	33	74
41	40	19	21	58
30	56	14	34	75
27	55	18	33	74
26	57	17	35	76
28	46	26	27	66
31	40	29	21	58
31	58	17	29	69
27	51	22	28	68
32	51	17	29	67

Figure 8. Neil’s recording sheet when drawing samples of size 100

Fish Caught from the dam			Interval estimate of percentage (%) of perch	
Bass(%)	Perch(%)	Trout(%)	lower limit	upper limit
28	55	17	34	76
41	40	19	27	58
30	56	14	36	78
27	55	18	34	76
26	57	17	37	79
28	46	26	30	66
31	40	29	27	58
31	52	17	31	73
27	51	22	30	72
32	51	17	30	72

Figure 9. Tim’s recording sheet when drawing samples of size 100

When Neil and Tim drew for sample #1 28% Bass, 55% Perch and 17% Trout, they reasoned as follows:

72. R: How will you calculate the first interval?

73. N: I’ve gone; I’ve given myself 20% on each side.

74. R: yeah.

75. N: I haven’t gone exact to 20% calculating for every fish. And each fish now is a whole percentage so these are more accurate.

76. T: Yeah.

77. R: So?

78. T: Sorta done the same. I’ve just gone on the 20% on each side, but not exactly. Just . . .

When forming estimate #1, both Neil and Tim decided to go about 20 above and below the observed value (55). First Neil explained that he had “gone 20% on each side” (line 73), and then Tim reiterated that he had done the same (line 75). The arbitrary nature of the value 20 is suggested by the fact that Neil chose to point out that he had not gone exactly 20 (line 75) and Tim reiterated “20% on each side but not exactly” (line 78). There was no explanation as to why the number 20 was chosen. Neil’s mathematical consideration of what each fish in the sample represented is clear in line 106 where he allocated “a whole percentage” to each fish, a conclusion he appears to have reached because there was now 100 fish in the sample. However, Neil does not appear to have necessarily used this information when making his estimate.

When Neil and Tim drew for sample #2 41% Bass, 40% Perch and 19% Trout, they reasoned:

- 79. T: I got 27 as my lower limit and 58 as the upper.
- 80. R: Ok. And you increased the lower limit by 3. Why did you increase it by 3? From 24 to 27.
- 81. T: Generally cause I thought it made more sense.
- 82. R: Is there any connection with those?
- 83. T: With this, I just looked at that and it’s 55, so that’s 15% more, so going down, by that much it sort of felt right. Same as that one.
- 84. N: My lower limit is 21 and my upper limit is 56.
- 85. R: And how did you find that?
- 86. N: It’s 40, which is 15 under 55, just gone down 15 on one side. . . enough to just go under which would be 19. Didn’t, I wasn’t didn’t or, yeah about 19. I didn’t think 19 was right nor was 21. So you just give an error, calculations. And then 40 equal 50, back off to 56.

When forming estimate #2, Neil and Tim used the difference between the observed value for sample #2 and the previously observed value (sample #1) to decide whether to increase or to decrease the previous estimate. As Neil explained “it’s 40 which is 15 under 55, just gone 15 on one side” (line 86). The difference (reduction of 15) between the observed values of 40 and 55 was used as a yardstick for the change in the estimate (a reduction) but ultimately the reduction was a little less than 15 because Neil explained that “I did not think that 19 is right” (line 86). Neil used a similar process for the upper limit including a fudge factor the effect of which can be seen in the change of his estimate from 56 to 58 in the upper limit estimate #2 (see Figure 8).

After drawing sample #3 they caught 30% Bass, 56% Perch and 14% Trout. Neil and Tim reasoned as follows:

- 87. N: My 34, upper limit is 75.
- 88. T: My lower was 36 and my upper was 78. So, it didn’t go up like, cause it’s only 1%. So I went up a little bit.
- 89. N: I guess I went up exactly 1 from our 55 answer.
- 90. R: Just 1?
- 91. N: Yeah, because I went up 1% more so it’s just by 1 off.
- 92. R: Ok, did you do the same?
- 93. T: Yeah.

When forming estimate #3 Tim used an approach already used with the smaller samples; he based his new estimate on a “convenient” previous estimate and adjusted up. In this case the observed value #3 percentage of Perch (56%) was very close in value to the first observed value #1 (55%) and Tim reasoned “only 1 %, so I went up a little bit” (line 88). Neil agreed that a change of 1 to form a new estimate was relevant (line 91).

After drawing sample #8 they caught 51% Bass, 52% Perch and 17% Trout. They reasoned as follows:

- 94. T: My lower was 34 and my upper was 71. I pretty much just looked at 55 and looked down from there.
- 95. R: You mean the first um, the first. . .
- 96. T: Um 34, 32 so. . . I went yeah; I went from 55 and notice the 3% difference.
- 97. R: So, here you just. . .
- 98. T: So yeah, so really it should actually be 31.
- 99. R: Subtract 3.
- 100. T: Yeah.
- 101. R: And what about the upper limit?
- 102. T: Umm, 3, 73. So I pretty much employed the same thing as what I did with the lower to the upper.

103. N: I've gone my lower limit of 29 and my upper limit as 69. That's because the lower 29 because I had 33 from 55 and for 46 I had 27. So it's roughly between those two. 3 down from both, 3 up and down from both of them roughly. So that was 29 and 74 and 66 and between them is roughly ... 69

By the time that Neil and Tim were forming estimate #8, they were using similar approaches to form their estimates. However, for estimate #8 they used different approaches. Tim used the approach where one "convenient" previously observed value is chosen, and the new estimate is formed by adjusting up or down. In this case the observed #1 (55) was chosen as the basis for the new estimate because it was close to the observed #8 (52) and the estimate #8 was made by taking 3 from each of the upper and lower limits in the old interval estimate (lines 94-102). On the other hand, Neil used the approach where two previously observed values are chosen that enclose the current observed value and then the limits of the new interval estimate are placed "between" the limits of the previous interval estimate. In this case, the observed value #8 of Perch (52) was between 55 (observed value #1) and 46 (observed value #6) and so Neil created an interval that was based on both estimate #1 and estimate #6. He selected the limits for the new interval so "it's roughly between those two". For the lower limit he explained how the estimate (29) was between 33 and 27, "3 down from both, 3 up and down from both of them roughly" (line 103).

7.4 Reflection Stage

In the final stage Neil and Tim were asked to reason about what had occurred in the first three stages by comparing the interval estimates of the percentage of Perch. During the first three stages the students were very focused on deciding on the numerical values for the interval estimates and provided only weak statistical reasoning in their explanations of how the values were chosen. During this stage the researcher tried to focus attention on the concepts of sampling, sample and variation as the students compared their estimates.

When asked to compare the estimates based on the samples of size 50 to those based on the samples of size 20 Neil reasoned:

104. R: If you will have a look at this one when we will draw a sample of 50 fish, and when we draw a sample of 20 fish, do you have the same intervals? Or do you have bigger or smaller intervals.

105. N: Mine are bigger because you know there's more fish and each percentage means more than when there was 20 fish... Of each percentage would have been 5.

Neil stated that "each percentage means more than when there was 20 fish" (line 105) and explained that 1% now represents 2 fish rather than 5 fish as it was for the sample size 20 situation. It is important to clearly mark here that Neil expressed himself wrongly: each fish is a higher percentage with a smaller sample size (e.g., with $n = 20$, 1 fish = 5%; with $n = 50$, 1 fish = 2%; with $n = 100$ 1 fish = 1%).

This alignment of a percentage of the sample representing actual fish was accurate but the way it was used to adjust the numbers for the upper and lower limits in the interval estimates indicates a lack of understanding of the difference between sample and population.

When asked to compare all three sets of interval estimates, i.e., for samples sizes 20, 50 and 100, Neil and Tim reasoned as follows:

106. T: Yeah. Generally when it went up, there's a lot more perch that were caught.

107. R: And what about the limit? The range, the interval.

108. N: The range is generally about, well, sometimes it's about 30, and I think nothing higher than 88.

109. T: Yeah, go from around that area.

110. N: There were some that were below that ... that are ... that would probably bring it down but generally about between 30 and 88. It wasn't too many below that. In the first few the lowest I got was 30, but apart from that they were all roughly in that range.

111. T: Yeah.

112. R: Ok, so and this range stays the same when you draw a sample of 20, 50 and 100 or not?

113. N: No, 20 do a lot lower numbers. ... That was 19, 13, and 28. So there was a lot more in the 30's and 24 ranges. Near low ones, but in the last ones we did, all my numbers were in the 30s.

114. R: Yeah.

115. N: 30s, there were umm 24 upwards pretty much. In the medium one, they were all similar. They did, they were building up. They have the lower class centre.

Initially in this reasoning Neil and Tim did not acknowledge that there were different size samples. They discussed the various interval estimates as if they were just in one big group of 30 interval estimates. When the researcher drew their attention (line 112) to the fact that the various estimates were based on different sized samples, Neil explained that the “20” samples had a “lot of lower numbers” (line 113) but he did not reason in a way that would indicate an appreciation of any impact that the smaller sample size might have had on the accuracy of the estimates.

As part of finding a possible range for the interval Neil (line 113) considered what was happening with the lower limits (about 30) and the upper limits (nothing higher than 88). Thus his interval became generally between about 30 and 88. However, when he realised that for one of the sample of 20 estimates he had written 13 as his lower limit, he clarified that sometimes the observed value was outside the 30 to 88 interval. Further discussion by Neil suggested that he was aware that the interval estimates he had created for the sample size 50 and sample size 100 situations were more consistent in terms of the numbers he had provided.

Generally a mathematical approach was taken to forming the estimates and little attention paid to statistical reasoning to explain the estimates. This provided interesting perspectives on the strategies used by the students to form interval estimates for the parameters but provided little information about the students’ conceptions of sampling, sample and variation. The following discussion outlines the estimation strategies used and the information that was gathered about the students’ conceptions of sampling, sample and variation.

8. Discussion

When forming interval estimates, the two secondary students generally focused on previously observed values to identify which previous interval estimates could be adjusted in some way to produce a new estimate. Although the students tended to experiment with choosing a “relevant” strategy each time a new sample (observed values) was drawn, a number of basic strategies emerged.

In spite of the apparently arbitrary choice of strategy to form estimates, the students had definite ideas about what they “expected”, as seen when they formed estimates but then decided to adjust them a little because what they “calculated” did not seem to be reasonable. This was evidenced in the use of terms such as “smudge” and “up a bit”, and also in the changing of values provided as estimates on the recording sheets. They were usually not able to articulate why they needed to make these small adjustments.

8.1 Strategies Used by the Students to Estimate

There were four main strategies, labeled S1 to S4, implemented by the students to reason when forming an estimate.

S1: only use the observed value. This involves no use of any information provided by previously observed values or previous estimates. This strategy was used by necessity when it was the first sample being drawn and was only used in such circumstances. The students made a random increase/decrease to the observed values #1 to form an estimate #1.

S2: choose one previous estimate and use it to form the new estimate. It was not always clear why a particular previous estimate was chosen in preference to others that were available. The increase/decrease made to the previous estimate, to form the new estimate, was dependent on the increase/decrease in the new observed value compared to the previously observed value. There were four different ways for deciding on the amount of the increase/decrease. The first (S2A) was to select an arbitrary value. For example, in his Sample Size 50 estimate #4 Neil talked about his upper limit going up by “a smudge”. Generally though there was some sense of proportionality so that if the increase/decrease in the observed value was small, then only a small increase/decrease was made in the interval limits. Similarly, large changes in the observed values led to large changes in the estimates. For example, in Neil’s Sample Size 50 estimate #3 he made the upper and lower limits “higher” because the observed value “was 4% higher”. In another example, after noticing that the observed value was lower than it ever was before, Neil reduced a previous estimate—“I’ve just pulled it back a bit”—and Tim reduced his estimate: “so I just bring it down”. There was no indication that the reduced amount was based on any mathematical calculation.

A second way (S2B) of deciding on the amount of increase/decrease involved using the difference between the current observed value and a previously observed value to determine the difference applied to the previous estimate to form the new estimate. Often a mathematical approach was taken to decide the size of the increase/decrease, or as stated by the students how much to go up and down. For example, Neil reasoned “it’s 40 which is 15 under 55, just gone 15 on one side”. The difference (reduction of 15) between the observed values of 40 and 55 was used as a yardstick for the change in the estimate (a reduction) but ultimately a little less than 15 because Neil explained that “I did not think that 19 is right” (first estimate of 33 minus 15 gives 19) and ultimately used 21 as the lower

limit. Neil used the same rule for the upper limit, including an adjustment (he said “a fudge”) that in fact made the reduction even more than 15. It is not clear why he chose to make the reduction even more rather than erring on the conservative side and making the reduction less than 15.

The third way (S2C) of deciding on the amount of increase/decrease involved a fixed amount being decided on depending on the sample size. For example, when the sample size was 50, each fish was 2%, and so increases or decreases of the percentage estimates were made in increments of “2%”. The problem with this processing is that the estimates being made were actually percentages as well and not actual numbers of fish, so the explanations were confusing at times.

The fourth way (S2D) of deciding on the amount of increase/decrease was to make a convenience choice of previously observed values that only differed slightly from the current observed value and thus meant little change was needed to produce a new estimate from the previous one. For example, Tim based his estimate #3, for Sample Size 100, on his estimate #1 because the difference between the new observed value (56%) and the first observed value (55%) was “only 1%, so I went up a little bit.” Another example is when Tim formed estimate #8 by reasoning that the observed (52) was 3 less than the observed value (55) for a previous estimate #1 and hence the new estimate (73) should be 3 less than the previous estimate (76).

S3: *choose two previous estimates and use them to form the required estimate.* Obviously, this strategy can only be used when there have been at least two previous estimates already made. The difference between the current and the two previously observed values is used to adjust the previous estimates in some way to form the new estimate. For example, two observed values were chosen that were either side of the current observed value, e.g., 40 and 50 were either side of “45”. Then the new estimate was chosen between the estimates for the two previously observed values, e.g., the lower limit 28 because “I’m going in between those two” (27 and 29). He chose the only number between 27 and 29. Again, for the upper limit “58” was chosen, which was between 51 and 65 and although it was the average it was not called that by the student who reasoned “I’ve just gone between those.” Similarly, in the Sample Size 50 estimate #5 Neil explained, “I just went between those two” (the observed 52 was between the previously observed values of 50 and 54).

S4: *choose a previous estimate and use it unadjusted.* This strategy was chosen when an observed value was exactly the same as a previously observed value. In this case the previous estimate was used as the new estimate. It was not considered necessary to form a new estimate or refine the estimate in any way. Thus it appears that the students ignored information from all other samples and also ignored any decision(s) already made with previous estimates.

Eventually, students appeared to settle on Strategy S3, which considered two previously observed values that were either side of the current observed value (e.g., 40 and 50 are either side of “45”) and then chose the estimate between the estimates for the two previously observed values.

Whilst creating the interval estimates, the students demonstrated very little evidence of the notion of stabilizing the values used for the limits of the interval estimate as more information becomes available with each successive sample. Hence, although the students looked back to previous estimates to actually form a new estimate, they had no sense of refining previous estimates. Also, although they did look to previous samples to help with the actual estimation, there was little evidence that they felt the need to be inclusive in any way of the previously observed values. The students did not appear to appreciate that many of the observed values for percentages of Perch should actually fall within the interval they are estimating.

The students did notice that the estimates that each makes are similar but see this as indicating they were “smart” and near the “right” answer rather than recognising that they were both trying to estimate the same parameter and so should be providing similar values.

8.2 Unusual Reasoning

There were six unusual aspects to the students’ reasoning. First, sometimes the interval limits were not consistently increased/decreased. For example, when the observed values went up the lower limit was increased but the upper limit was decreased. So effectively the interval was made wider rather than simply sliding so that interval width was maintained but the centre was changed.

Second, sometimes intervals were proposed that were not inclusive of the observed values. For example, in Sample Size 20 estimate #7, the upper limit of the interval was set at a value that was lower than five of the observed values but no concern was shown by the students that these observed values were excluded from the proposed interval

estimate.

Third, the samples drawn were generally considered to be the starting point for estimation. For example, Neil's estimate #10 used the term "starting number" to refer to the observed number of perch on which the estimate was to be based. This suggested that he viewed the observed number of perch as the first number he should consider when making an estimate. In fact, he actually called the entire column of observed values "starting numbers".

Fourth, there was some evidence that students had intuitive expectations of what would be observed. For example, when one sample had a very low value for perch compared to other samples, the students began speculating on the reasons for such value and clearly showing that they had some expectation of what were typical observed values.

Fifth, similarity in estimates formed was considered by the two students as confirming that their estimate was close to the "correct" value. By the time the students were creating Sample Size 50 estimate #8, they had provided almost identical estimates and Neil noticed that the estimates were similar—"our numbers are getting very close"—but when asked to explain why this might be, he said, "just getting very similar," and jokingly added, "we are both smart guys, it must be the answer." In another instance, Neil referred to being close to the "correct answer" when he and Tim had the same estimate for the upper limit of the interval.

Sixth, a direct mathematical representation was assigned to each percent of fish caught and often used to calculate the change in interval estimates. When asked to compare the estimates based on the samples of Size 50 to those based on the samples of Size 20, Neil said that, "each percentage means more," meaning that each fish is a higher percentage with a smaller sample size (e.g., with $n = 20$, 1 fish = 5%; with $n = 50$, 1 fish = 2%).

A possible interpretation of his statement is that the sample of Size 50 provides more information for making an estimate but this is obviously not the case from his explanation. Then, early on with the Sample Size 100 activity Neil explained that, "each fish now is a whole percentage so these are more accurate". This is a reference to the fact that 100 fish are caught, so each percentage represented one fish when adjusting estimates. The students gave the impression that this was going to make everything simpler mathematically when creating estimates.

8.3 Evidence of Understanding of Core Concepts (Sampling, Sample, Variation)

When making and explaining interval estimates, the students' expressions did not indicate any evidence of their understanding of the concept of samples or of sampling. For example, it was expected that students might have chosen to discuss how the fish were selected (the process of sampling) or whether the particular group of fish chosen (the sample) was representative of the population of fish in the dam, but this was not the case. Also, it was expected that students might have discussed a connection between the size of the sample and the width of the interval estimates they had created.

There was some evidence of the students' understanding of variation but no explicit use the terms variation, vary, or varying. There were five different ways that some understanding of variation was implied when the students were sampling in the *Sample Size 20 Stage* and *Sample Size 50 Stage*. First, Neil explained the upper limit of one of his interval estimates by stating, "'cause it needs a really high number there". This implies that Neil had some expectation of how large the number should be and thus of how much variation there should be.

Second, Neil explained that his interval estimate "correlates with my earlier percentages". The students had not studied the statistical concept of correlation and Neil used the non-mathematical meaning of the word, that two things are related and connected to each other. This implies that Neil was considering how the interval estimates were changing. The term "percentages" was used because the actual data was being expressed as percentages.

Third, Neil explained that the data had "repeating samples" when the numbers in the sample were the same as for a sample drawn earlier. This implies some appreciation of the fact that the samples will vary, although focusing on the fact that the same sample had previously occurred may.

Fourth, Neil chose his interval estimate "keeping room for error". This implies that he has recognized that the samples are varying and he needs to allow for this.

Fifth, Tim "tried to balance it out a little bit" when deciding on the values at either end of his interval. He provided a similar expression about balancing or choosing values so they balance for many of his estimates. This implies he was considering how the limits of his interval should be spread either side of a value but he does not indicate the value itself. Similarly, when forming estimate #2, Neil chose numbers either side of the previously observed value ("both numbers 20 outside"), giving some indication of consideration of variability. Then during the *Sample Size 100 Stage*, there was another expression, "20% on each side", introduced by Tim that continued the theme of

balancing. Neil then responded “either side getting there giving room for error”, which continued Tim’s theme of either side/balance but also implied that he was trying to resolve this against his own notion of allowing room for error.

9. Limitations of the Study

There were one main limitation to this study, related to task design, sample size, and interview technique that prevent the results from being more widely applicable. First, the task was designed so that the requested estimation was for a percentage (of perch in the dam) and there were times when it was not clear that the students understood the difference between actual number of fish and the percentage of fish. A task that requires interval estimates of actual numbers, rather than percentages, may prove to be simpler for students when forming interval estimates for a parameter of a population.

10. Future Research

The above research suggests some aspects of students’ reasoning while making interval estimates. While interesting estimation strategies emerged, there was little that was evidenced in relation to students’ conceptions of the core concepts, sampling, sample and variation. The results of this research study provide those teaching estimation to secondary students with better insights into their students’ reasoning.

In raising the idea that there may be a better way to examine statistical reasoning, particularly as involved in the sampling process, it is acknowledged that the focus should not be solely on mathematical approaches to estimation. There is fascinating research to be done in investigating students’ statistical reasoning when sampling and making interval estimates. There are however important research questions that the above research has not addressed. Two important issues for further research arise out of this study. The first relates to the importance of engaging students in quantifying the level of confidence (Ben-Zvi et al., 2012; Prodromou, 2011) when making informal statistical inferences about samples and sampling. The estimation tasks used in the reported study could be expanded to include student expression of their level of confidence in relation to their informal inferential reasoning while sampling. While developing an interval that includes possible values for a particular parameter is important, it is also relevant to be aware of the level of confidence (certainty) in relation to the interval estimate. Thus, there are two aspects of using intervals to estimate parameters, one is the *interval estimate* itself and the other is the *level of confidence* of the number of times that, with repeated sampling, this interval “will” contain the true value of the population parameter. Students engaged in initial reasoning cannot be expected to formalize their level of confidence in their estimates, in the same way that formal statistics would use confidence intervals, so further research into the students’ level of confidence might be pursued. This could be achieved, for example, by not setting the number of samples as “10” and letting the students sample until they are confident that they form a “good” interval estimate, i.e., students decide when to stop sampling. The second issue for further research is a need to investigate how the instructional idea of growing samples (Ben-Zvi et al., 2012) can be further improved and used in combination with another instructional idea, namely “Shrinking samples”. Insights into this instructional approach may provide a useful perspective of the role of such activities in the development of students’ reasoning about sampling, samples, and variation.

References

- Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal*, 3(2), 64-83.
- Bakker, A., Kent, P., Derry, J., Noss, R., & Hoyles, C. (2008). Statistical inference at work: Statistical process control as an example. *Statistics Education Research Journal*, 7(2), 131-146.
- Bakker, A., & Derry, J. (2011). Lessons from inferentialism for statistics education. *Mathematical Thinking and Learning*, 13(1&2), 5-26. <http://dx.doi.org/10.1080/10986065.2011.538293>
- Ben-Zvi, D., Aridor, K., Makar, K., & Bakker, A. (2012). Students’ emergent articulations of uncertainty while making informal statistical inferences. *International Journal on Mathematics Education (ZDM)*, 44(7), 913-925. <http://dx.doi.org/10.1007/s11858-012-0420-3>
- Camtasia: Tec Smith Corporation. (2000). *Camtasia studio* (Accessed 10 March 2013, Version 6.0) [Computer software]. Okemos, MI: Tec Smith Corporation. Retrieved from <http://www.techsmith.com/camtasia.asp>
- Chance, B., delMas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi, & J. Garfield (Eds.), *The Challenge of developing statistical literacy, reasoning, and thinking* (pp. 295-323).

Boston: Kluwer Academics Publishers.

- Farlex. (2012). *The Free Dictionary* (Accessed 10 March 2013). Retrieved from <http://www.thefreedictionary.com/Dictionary.htm>
- Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. New York: Springer.
- Konold, C., & Pollatsek, A. (2002). Data analysis as a search for signals in noisy processes. *Journal for Research in Mathematics Education*, 33(4), 259-289. <http://dx.doi.org/10.2307/749741>
- Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, 8(1), 82-105.
- Pfannkuch, M. (2005). Probability and statistical inference: How can teachers enable learners to make the connection. In G. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 267-294). Dordrecht, the Netherlands: Kluwer Academic Publishers. http://dx.doi.org/10.1007/0-387-24530-8_12
- Pfannkuch, M. (2008). Building sampling concepts for statistical inference: A case study (Accessed 10 March 2013). Paper presented at the 11th International Congress on Mathematic Education, Monterrey, Mexico. Retrieved from <http://tsg.icme11.org/document/get/476>
- Pratt, D., & Ainley, J. (Eds.) (2008). Informal inferential reasoning (Accessed 10 March 2013). Special Issue of the *Statistics Education Research Journal*, 7(2), 3-4. Retrieved from [http://www.stat.auckland.ac.nz/~iase/serj/SERJ7\(2\)_Pratt_Ainley.pdf](http://www.stat.auckland.ac.nz/~iase/serj/SERJ7(2)_Pratt_Ainley.pdf)
- Prodromou, T. (2011). *Students' emerging inferential reasoning about samples and sampling* (Accessed 10 March 2013). In J. Clark, B. Kissane, J. Musley, T. Spencer, & S. Thornton (Eds.), *Proceedings of the 34th Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 640-648). Alice Springs, Australia: MERGA34. Retrieved from http://www.merga.net.au/documents/RP_PRODROMOU_MERGA34-AAMT.pdf
- Robson, C. (1993). *Real World Research*. Oxford: Blackwell.
- Rossman, A. (2008). Reasoning about informal statistical inference: One statistician's view (Accessed 10 March 2013). *Statistics Education Research Journal*, 7(2), 5-19. Retrieved from [http://www.stat.auckland.ac.nz/~iase/serj/SERJ7\(2\)_Rossman.pdf](http://www.stat.auckland.ac.nz/~iase/serj/SERJ7(2)_Rossman.pdf)
- Watson, J. M. (2004). Developing reasoning about samples. In D. Ben-Zvi, & J. Garfield (Eds.), *The challenge of developing g statistical literacy, reasoning and thinking* (pp. 277-294). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Watson, J. M. (2006). *Statistical literacy at school: Growth and goals*. London: Lawrence Erlbaum Associates.
- Watson, J. M., & Moritz, J. B. (2000a). Developing concepts of sampling. *Journal for Research in Mathematics Education*, 31, 44-70. <http://dx.doi.org/10.2307/749819>
- Watson, J. M., & Moritz, J. B. (2000b). Development of understanding of sampling for statistical literacy. *Journal of Mathematical Behavior*, 19, 109-136. [http://dx.doi.org/10.1016/S0732-3123\(00\)00039-0](http://dx.doi.org/10.1016/S0732-3123(00)00039-0)
- Wild, C. (2006). The concept of distribution. *Statistics Education Research Journal*, 5(2), 10-26.
- Zieffler, A., Garfield, J., delMas, R., & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, 7(2), 40-58.