World Academy of Science, Engineering and Technology
International Journal of Educational and Pedagogical Sciences
Vol:7, No:6, 2013

Informal Inferential Reasoning Using a Modelling Approach within a Computer-Based Simulation

Theodosia Prodromou

***Abstract*—**The article investigates how 14- to 15- year-olds build informal conceptions of inferential statistics as they engage in a modelling process and build their own computer simulations with dynamic statistical software. This study proposes four primary phases of informal inferential reasoning for the students in the statistical modeling and simulation process. Findings show shifts in the conceptual structures across the four phases and point to the potential of all of these phases for fostering the development of students' robust knowledge of the logic of inference when using computer based simulations to model and investigate statistical questions.

***Keywords*—**Inferential reasoning, learning, modelling, statistical inference, simulation.

I. Introduction

STATISTICS is becoming increasingly important to all levels of citizenship, with an abundance of data available to inform decision-making. A solid understanding of *inferential statistics* is of major importance for designing and interpreting empirical results in all scientific disciplines. This topic of statistical inference is relevant for the development of research in all empirical sciences, including psychology and education as well [5]. Statistical inference receives a particular attention in courses of statistics where hypothesis test and confidence intervals are taught to students as the method for analysis data to evaluate scientific hypotheses (see [1]). However, students are usually prone to fall into many misconceptions when making statistical inferences (see [10]) because inferential statistics involve understanding of many abstract concepts such as sampling distributions and significance level. The concept of sampling distribution is generally poorly understood as discussed elsewhere [6], [12]. Many students often are unable to integrate the different ideas pertaining inferential reasoning and use concepts in inferential reasoning [3]. Students often held misconceptions about the law of small numbers, sampling variability, sample mean and the properties of its sampling distribution such as the law of large numbers. In particular many students neglected the effect of sample size on the variance of the sample mean [6]. Students often confuse the *population* and the *sampling* distributions [6] and might not be able to identify the difference between the distribution of a *sample* and the *sampling* distribution of a statistic [12]. Moreover, students seemed to be confused when they try to justify the use of the Normal distribution, although they might be comfortable doing the formal manipulations needed to use the central limit theorem (see [20]). As discussed by Batanero, Tauber and Sanchez [4] students have difficulties in distinguishing between the real sampling distribution and the theoretical model of a normally distributed population that is used as an approximation of the sampling distribution, based on the central limit theorem that is used to test the null hypothesis in a significance test. Further research on students' statistical inferential reasoning shows that students have many difficulties in connecting the available evidence with the question under investigation to draw inferences [14]. Other learning difficulties are related to building a schema of interrelated statistical concepts, such as distribution, sampling variability, and representativeness (see [17], [19]).

This paper, instead of dismissing or eradicating misconceptions, we should consider them as starting points which provide a pedagogic challenge of how to build on learners' impoverished view of statistical inference and help students develop effective secondary (scientifically learned or taught) intuitions of inferential statistics. The following section examines statistical inferences and learning as advocated by statistics educators.

T. Prodromou is with University of New England, Armidale, NSW 2350 Australia (phone: +61 2 6773 3237; e-mail: theodosia.prodromou@une.edu.au).

II. Theoretical Framework

Before students are introduced to statistical inference methods to decide whether the patterns they observe in data are real or random, they are usually presented with statistical problems that require informal inference.

Informal Inferential Reasoning is the process of drawing generalized conclusions from data. Four critical principles have been identified as important to making informal inferences from data: (1) generalizing beyond data (parameter estimates, conclusions and predictions); (2) using data as evidence of the generalization; (3) articulating the degree of certainty (due to variability) embedded in the generalization (these three principles were articulated by [13]); and (4) comparing datasets with a model such as ideal (targeted) distributions (proposed by [2]). 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.

Konold and Pollatsek [11] recommended that early teaching of statistics should focus on informal methods of data analysis. Many statistics educators advocate that inference should be taught entirely from an empirical perspective through simulation methods that enable students to better think statistically (as discussed elsewhere [9], [7]). The empirical study of sampling variability typically focuses on drawing repeated samples from a population, forming a distribution of sample statistics (such as sample means) from those repeated samples, and comparing the observed sample statistic to the empirical sampling distribution. This resampling approach elucidates how probability provides a theoretical structure for statistical inference, as it is based on the notion of considering what would happen when comparing the observed sample statistic to the distribution of sample statistics created under a chance model. As argued by Phannkuch [14] "the resampling approach to teaching would appear to be the most promising direction, as it could enable students to link probability intuitively with statistical inference" (p. 290). For such an approach, "randomization-based inference makes a direct connection between data production and the logic of inference that deserves to be at the core of introductory statistics" (see [7], p. 130).

According to Cobb [7] the core logic of inference entails the 3Rs: (a) model set up that allows for randomize data production; (b) repeat by simulation to generate data from (a model) for a single trial and assess whether the outcomes are reasonable. Specify the summary measure to be collected for each trial and generate data for many trials collecting the summary measure each time; (c) Reject any model that does not accurately represent the phenomenon it was intended to model., Cobb suggests that digital technologies provide a natural way to introduce students to computer-intensive simulation-based methods, allowing students to easily create simulation models, and then to interpret the observed outcomes.

This change of focus in statistics pedagogy-from centering on computation (i.e., the normal distribution and procedures, formulaic hypothesis tests), to the core logic of inference (i.e., chance models, and determining statistical unusualness; [7] has led to some reconceptualization of teaching statistics.

Inspired by Cobb [7] new curricula have been emerged in the last two years. Such curricula focus predominantly on using ideas of chance and models, along with computer simulations (i.e., Chance Agents for Teaching and Learning Statistics [CATALST], [8]) and randomization-based techniques, to make and understand statistical inferences. CATALST immerses students in the simulation-based approach to statistical inference that requires students to create a model with respect to a specific context, repeatedly simulate data from the model, and then use the resulting distribution of a particular computed statistic to draw statistical inferences.

Efforts should be made to fully assess the pedagogical value of computer-intensive and simulation-based methods when teaching the logic of inference, and to investigate the linking of ideas of variation and probability. Such novel methods inevitably bring with them new challenges in how students learn and give rise to research questions about the conceptual development of students who engage in constructing such chance models. It is important to understand how students construct models, run simulations and interpret outcomes, and reason about uncertainty in the context of making informal statistical inferences, as well as understanding the challenges students might encounter in such a pedagogical approach.

This paper presents data from a study in Australian schools, focusing on how Grade 9 students develop informal conceptions of inferential statistics as they engage in modelling using TinkerPlots2 [18] computer-based simulations. It is noteworthy to point out that although the terms "randomization" techniques and the "chance variation" to the extent that it relies on formal probability are precluded from secondary school curricula, this paper introduces randomization to Year 9 students through simulation-based approaches. Moreover, chance variation as an idea is assessed in an intuitive fashion.

III. Methodology

Thirty students in Grade 9, ranging from 14 to 15 years in age, from a rural secondary school in New South Wales, Australia, formed the population of this study. The researcher spent 2 sessions (40-45 minutes each) introducing the class teacher and the students to *Tinkerplots2* during regular mathematics lessons. All students were familiarised with the *TinkerPlots2* software, explicitly focusing on learning skills related to *TinkerPlots2*. In the first session, all students watched instructional movies that show how to use *TinkerPlots2* features to build a simulation. In the second session, all students were familiarized with *Tinkerplots2* through a number of introductory activities related to building a data factory that simulates real phenomena. The students also ran a simulation and observed the generation of data, and the distributions of the various data.

Ten average-ability students volunteered to spend a third session, outside of class time, to engage in the task reported in this study. In this session, students were asked to use the tools of *TinkerPlots2* and the "Data Factory" features (see Fig. 1) to generate a simulation to investigate the impact of hours spent using Facebook on the school performance of a group of students. The students created a number of "virtual students," each defined by several variables (gender, hours spent on Facebook per week, school performance) whose values were determined by *Tinkerplots2* using pre-defined probability distributions. After constructing their model, students were asked to run the simulation and interpret the outcomes.

Each session lasted approximately 45-60 minutes and each pair of students worked directly with the researcher. The researcher interacted continuously with the students in order to observe the reasoning they used to explain the data and simulations. The data collected included audio recordings of each pair's voices and video recordings of the screen output on the computer activity using Camtasia software.

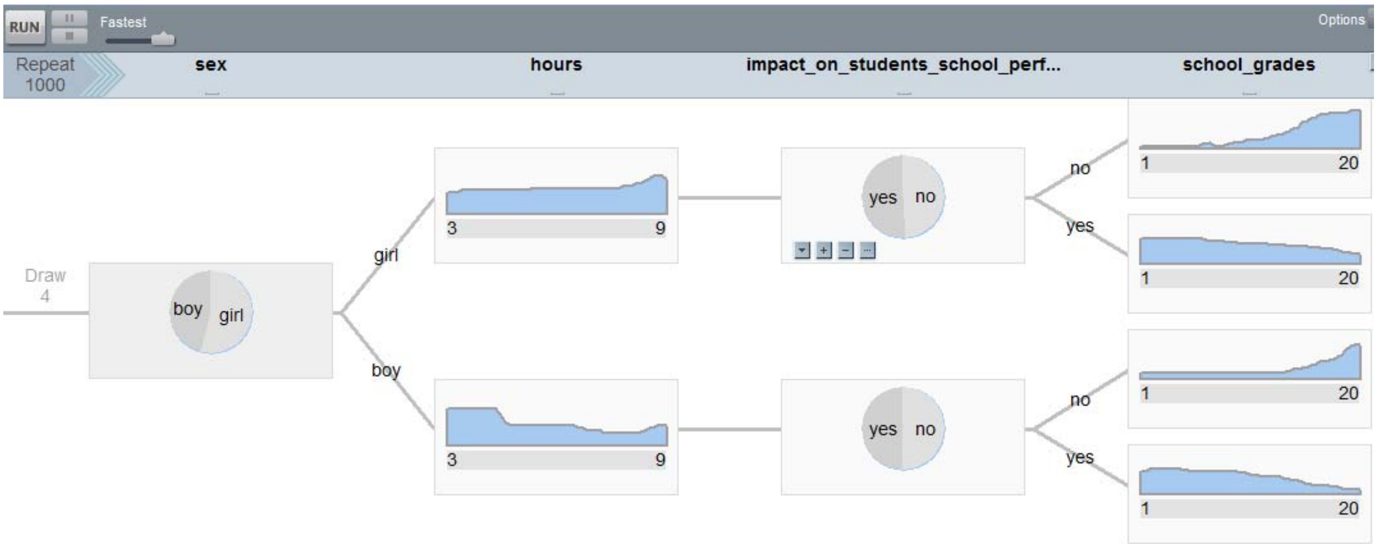

Fig. 1 Data Factory that simulates the Facebook Task

In tune with Cobb's call (see [7]) for teaching the logic of inference and the related theoretical framework to investigate how students might construct models, run simulations, and interpret outcomes. First, the audio recordings were transcribed and screenshots were incorporated as necessary to make sense of the transcription.

The data were then analysed using progressive focusing (see [16]), a process by which the author began with a wide field of focus and gradually narrowed the field by identifying key foci for ensuing study.

This article focuses on one pair of students, George (G) and Rafael (Ra). Although the same insights as reported below were evident in the analysis of the sessions of other pairs of students, George and Rafael provided (in my view) the clearest illustration of how students how students construct models, run simulations, interpret outcomes, reason when making informal statistical inferences.

IV. Results

The session began with George and Rafael creating a model to generate data to reasonably simulate the Facebook problem (Fig 1). The students began their attempt by perceiving a holistic entity, such as a student, as consisting of a cluster of pieces of data having attributes such as gender, hours spent on Facebook, impact of hours spent on FB on students' performance, and school grades (Fig. 1). When the students drew curves in the *Tinkerplots2* interface to define the probability density functions that would be used to generate the simulation data, they appeared to use software interface that relies on signal, variation, and spread of data to create the model. They talked about the curves with respect to most common values (what we could consider the "signal"), and variation of these values (what we could consider the "noise") [15].

After the boys had generated 1000 virtual students, they looked at the distributions created in the sampler and the distributions of generated data.

1. G: Okay, so these areas here (pointing to the circled areas of the graph (bottom) of Fig. 2). I reckon, they're just spots where, they could just be ugh, smaller populations just happening to do that. It's just, it's hard to explain.
2. Re (Researcher): What do you mean?
3. G:See, on the graph that we've put here … there's no spike here (pointing to the graph (top) of Fig. 2) and yet here there are these spikes (pointing to the circled areas of the graph (bottom) of Fig. 2)) and um, I just think that they're just people who happened to go on for 6 hours.
4. Ra: Yeah… They probably had more free time… or cold weather.
5. G: And, there are more people doing that say 6 hours but on the graph (top Fig. 2) it doesn't show that.

The students' attention was attracted by slices of prominent features of the observed distribution of hours spent per week by males (see bottom Fig. 2), such as higher areas of accumulated data compared to the curve that defined a probability density function of the hours spent on FB by males (see top Fig. 2). George seemed to refer to the variation caused by small samples when he referred to "the smaller populations" (line 1). Line 1 indicates that George seemed to recognize that the small sample size introduced variation, but he was unable to explain such vagaries of variation. They attempted to attribute some common cause factors to normal day-to-day variation, for example, people who happened to stay on for 6 hours (line 3), who have free time (line 4), or because of cold weather (line 4). When the students compared the outcome data distribution of the hours spent per week by students (Fig. 2) on FB to the distribution they created in the sampler, they began considering the possible outliers and other interesting features of the observed data. In particular, they reasoned about the data in terms of possible outliers, clusters of data and interesting individual cases.

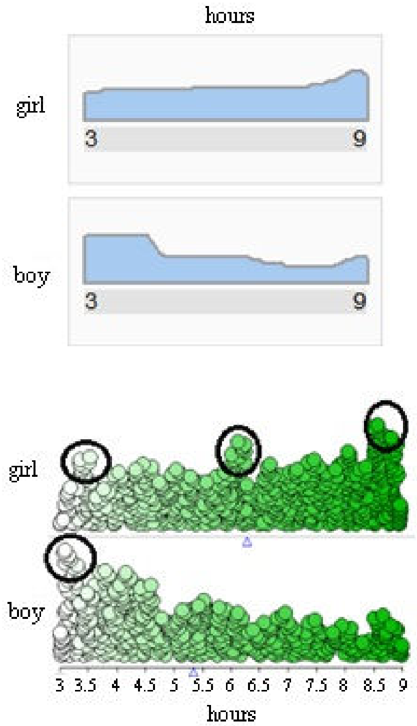

Fig. 2 Distributions created in the sampler (top) and distributions of generated data (bottom)

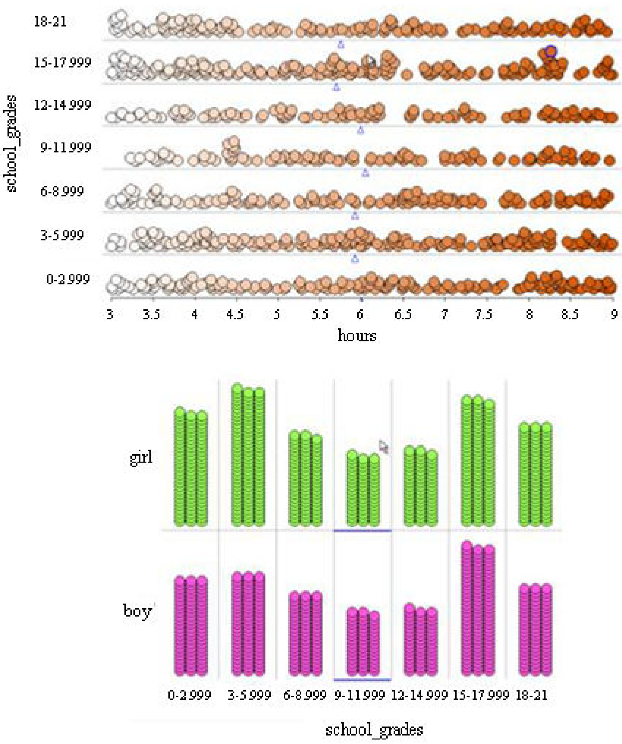

Fig. 3 Distributions of generated data (top) and distributions of school grades versus gender (bottom)

6. Re: What do these graphs show? Which graph shows the actual results?
7. G: These are the actual results, up here … and it's been put into that because that's how what we think.
8. Ra: That's the results that just come out.
9. Re: Do you think the results show what you think?
10. G+Ra: More or less.
11. G: They could not always be the same. Maybe, sometimes they will be the same.
12. Ra: When we will observe the hours spent on FB by a really big sample of students, like 50000 students or even more…

The above quotes show that students used the model to generate a single trial of the experiment and they investigated the outcomes from a single trial. George appeared to distinguish the distribution of the actual results from the model they created when they drew the density function. He very eloquently articulated that the chance model presented by the density function as drawn in the sampler shows the "model" they constructed in their mind (line 7). He seemed to have a vague sense of the actual results being generated by this model when he mentioned that "it's been put into that". Both boys articulated colloquial notions of chance when they used expressions such as "more or less", and "maybe" considering the appropriate phrasing of the statistical questions, that can be attributed to their preliminary sense of chance or variability between different outcomes.

The students expected that the distribution of the generated data would resample sometimes the model (lines 11). George was uncertain about the absolute resemblance of the empirically observed distribution of hours spent on FB to the curves students created in the sampler. One possible interpretation of this situation, could be that when the model is run a few times to simulate a number of students, there is stability in the peaked data but there is some variation observed in the general details of the shape. On the contrary, Raphael seemed to coordinate the role of sample size when the number of simulated students would increase (Law of Large Numbers) and the pattern would became more obvious in the variability within the data set.

When they observed the distribution of generated data (see Fig. 3 (top)):

13. G:It doesn't really seem to have an effect on how much you use it (referring to FB). The grades are still more or less the same.
14. Ra: Yet there's no one there and there are gaps.
15. G: There are gaps, like here, and there's gaps here...Okay, there's a gap there. Then there's a big bulk here. The bulks, but there's still individual like here. And there's an individual… (while speaking he points to places on the graph, see Fig. 3 (top))
16. Ra: And they just stand out, there's nothing around them. And there's big clusters of people, where there's circles overlapping other circles, and like here (pointing to the line 0-2.999 of Fig. 3 (top)) and here, everywhere.

17. G: It cannot be similar to this (pointing to the graph of school grades in Fig. 1). It is confusing.
18. Ra: And they just stand out, there's nothing around them. And there's big clusters of people, where there's circles overlapping other circles, and like here and here, everywhere.

The above quotes show George and Raphael when they used the model to simulate data for many trials (each time interpreting the results). The students examined the empirically observed distributions of the resulting outcomes. They used the observed distribution to assess particular outcomes and compared the behaviour of the model to the observed data. At first sight, it seemed puzzling that the boys did not 'see' the impact of hours (spent on FB) on students' school grades. However, their extended discussion showed how their attention was, at this point focused on interesting individual cases (lines 16), areas of the graphs that there were no data values and areas where there were big clusters of data values. The boys seemed unable to explain the relation between school's grades versus hours spent on FB. When George attempted to compare the observed graphs to the density function drawn in the sampler, he found the information displayed in the graph confusing (line 17). He could not even observe that when the model was run a few times, there was stability in the peaks but there was some variation observed in the general details of the shape. When the boys observed the distribution of school grades for each gender:

19. G: Girls either averaged like really well, or not very well but boys are sort a just the same.
20. Ra: But, there also seems to be a lot of boys that averaged very well.
21. G: Fairly well, then there's a couple that do very well. But there are less numbers than this, 9 to 14, 9 to 15 range of um school grades. And that seems to be the same with girls as well
22. Re: Do you think that, these are the results what you expected?
23. Ra: I reckon girls should've had a bigger bulge just up around there (pointing to the range of 15-20 on the graph of Fig. 3(bottom)).

Raphael was expecting to observe an increase in the school grades of the female population (line 23). The boys then suggested to change the model that simulates the empirical grades for girls, increasing the grades for girls. They looked at their previous models and also suggested redrawing the density function that generates the boys' school grades:

24. G: Make it go down (referring to the curve of the density function of boys' school grades) Make it go down, not as steeply like that.

Students revisited their previous actions and co-ordinated associated actions and generalisations to make adjustments to the models redrawing the density functions in the sampler that generate the empirical data.

V. Discussion

The results suggest that both the modelling process and the simulation process appear to be appropriate resources for introducing beginning inference to middle school students. The modelling and simulation process challenged students to construct models, interpret empirically observed distributions, compare the behaviour of the models to empirically observed data and evaluate the models used to generate data. Computer-intensive modelling and simulation-based methods reinforce each other, explicitly in terms of the elements of understanding the constructing models (samplers) appropriately to model the statistical problem, generation of simulated data, examination of the empirically observed distribution of the observed outcomes, interpretation of the results, and evaluation of the model used to generate empirical data. Thus these data, though only suggestive, seem to indicate a path by which Cobb's conjecture regarding the value of digital technologies in statistics education can be elaborated.

Focusing on Cobb's [7] logic of inference, the results of this preliminary research illustrate the conceptual structures that students build about informal inferential reasoning across four phases that trace the movement of students' informal inferential reasoning in the modelling and simulation process (Table I).

TABLE I
Phases of Informal Inferential Reasoning in the Modelling and Simulation Process

Phases	Description
Phase 1	Specify a model that will generate data to simulate the experiment. Use software interface that relies on signal, variation, and spread of data to create the model.
Phase 2	Use the model to generate a single trial of the experiment, investigate the outcomes from a single trial; Construct an appropriate representation of the outcomes from the single trial; interpret the results of a single trial; Consider possible outliers and other interesting individual cases.
Phase 3	Use the model to generate simulated data for many trials, each time interpreting the results. Examine the empirically observed distributions of the resulting outcomes. Use the observed distribution to assess particular outcomes. Compare the behaviour of the model to observed data; evaluate the model; consider the Law of Large Numbers when developing understandings between observations of distributions of empirical data and the model.
Phase 4	Coordinate the actions of phases 1-3 to change the model, interpret the results, draw inferences based on the data at hand.

The phases need to be refined before trying to apply them to a pedagogical practice.

In terms of assessment, combining the four phases of informal inferential reasoning with a developmental model from cognitive psychology might be promising in documenting students' continuous progress of reasoning. As such, the researcher assumes it would shed some light on the structural complexity of making informal statistical inferences.

Given the observation of students' initial difficulties in creating an appropriate model that would generate empirical data, it would appear that the use of a dynamic software such as Tinkerplots2 may contribute to developing intuitions about what might be considered "appropriate co-ordination of signal and noise" or "meaningful approximation of real or simulated phenomena" [15].

The four phases describe the conceptual structures that students build about informal inferential reasoning and provide a base to trace the cognitive processes involved in inferential reasoning. Future research will investigate more systematically students' phases of informal inferential reasoning in the modelling and simulation process.

REFERENCES

[1] Aberson, C. L., Berger, D. E., Healy, M. R., & Romero, V. L. (2003). Evaluation of an interactive tutorial for teaching hypothesis testing concepts. *Teaching of Psychology, 30*, 76-79.

[2] 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. http://www.stat.auckland.ac.nz/ iase/serj/SERJ7(2)_Bakker.pdf. Accessed 10 January 2013.

[3] Batanero, C. (2005). Statistics education as a field for research and practice. In *Proceedings of the 10th international commission for mathematical instruction*. Copenhagen, Denmark: International Commission for Mathematical Instruction.

[4] Batanero, C., Tauber, L. M., & S´anchez, V. (2004). Students' reasoning about the normal distribution. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 257–276). Dordrecht, The Netherlands: Kluwer Academic Publishers.

[5] Belia, S., Fidler, F., Williams, J., & Cumming, G. (2005). Researchers misunderstand intervals and standard errors bars. *Psychological Methods, 10, 389-396.*

[6] Chance, B., delMas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi and J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking* (pp. 295-323). Dordrecht, The Netherlands: Kluwer Academic Publishers.

[7] Cobb, G. W. (2007). The introductory statistics course: a ptolemaic curriculum? *Technology innovations in statistics Education,* 1 (1).

[8] Garfield, J., delMas, B., & Zieffler, A. (2012). Developing statistical modelers and thinkers in an introductory, tertiary-level statistics course. *ZDM−The International Journal on Mathematics Education*, *44*(3), 883-898. doi:10.1007/s11858-012-0447-5

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

[10] Kirk, R. E. (2001). Promoting good statistical practices: Some suggestions. *Educational and Psychological Measurement,* 61 (2), 213−218.

[11] Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education, 33*(4), 259-289.

[12] Lipson, K. (2002). The role of computer based technology in developing understanding of the concept of sampling distribution. In *Proceedings of_the sixth international conference on teaching statistics*. Voorburg, The Netherlands: International Statistical Institute.

[13] Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal, 8*(1), 82-105. http://www.stat.auckland.ac.nz/ iase/ serj/SERJ8(1)_Makar_Rubin.pdf. Accessed 10 January 2013.

[14] 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. 267-297). Dorcdrecht, The Netherlands: Kluwer Academic Publishers.

[15] Prodromou, T. (2012). Developing a modeling approach to probability using computer-based simulations. *Proceedings of the 12th International Congress on Mathematical Education*, Topic Study Group 11, pp. 2394-2403. COEX, Seoul, Korea: ICME. http://www.icme12.org /upload /UpFile2/TSG/0978.pdf. Accessed 10 January 2013.

[16] Robson, C. (1993). *Real world research*. Oxford, England: Blackwell.

[17] Saldanha, L., & Thompson, P. (2003). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics*, 51, 257-270.

[18] *TinkerPlots: Dynamic data exploration* (Version 2.0) [Computer software]. Emeryville: CA: Key Curriculum Press.

[19] Watson, J. M., & Moritz, J. B. (2000). Development of understanding of sampling for statistical literacy. *Journal of Mathematical Behaviour, 19,* 109-136.

[20] Wilensky, U. (1997). What is normal anyway? Therapy for epistemological anxiety. *Educational Studies in Mathematics*, *33*(2), 171–202.