A FRAMEWORK FOR ANALYZING INFORMAL INFERENTIAL REASONING TASKS IN MIDDLE SCHOOL TEXTBOOKS

Maryann E. Huey
Drake University
maryann.huey@drake.edu

Christa D. Jackson
Iowa State University
jacksonc@iastate.edu

Building upon the work of Zieffler, Garfield, DelMas and Reading (2008) and others, we developed a framework for assessing informal inferential tasks in middle school mathematics textbooks. The framework both embodies the key recommendations for developing informal inferential reasoning and captures common trimming attributes, which lower the cognitive demand and opportunities to learn. Researchers believe that introducing inferential reasoning informally will assist students later in developing argumentation structures necessary for understanding formal methods (Wild & Pfannkuch, 1999). Inferential reasoning has long been a key learning goal of statistics education and provides access to viewing knowledge of core statistical concepts and reasoning about data distributions. Tools are needed to assess the fidelity of tasks in alignment with both national and research-based recommendations.

Keywords: Curriculum Analysis; Data Analysis and Statistics; Middle School Education

Background

Inferential reasoning has served as a unifying theme and goal of introductory statistics courses at the tertiary level for a number of years (Konold & Pollatsek, 2002). With the recent emphasis of statistics as a core component of the middle and secondary mathematics curriculum, the role of inference is gaining in prominence (NGA Center & CCSSO, 2010). Current recommendations for middle and secondary statistics education outlined in the *Guidelines for Assessment and Instruction in Statistics Education [GAISE]* report support the introduction of informal inferential reasoning at the middle school level and formalizing inferential reasoning during the secondary years (Franklin et al., 2007). These recommendations are evident in the articulation of the Common Core State Standards for Mathematics (CCSS-M) adopted throughout the United States (NGA & CCSS0, 2010), but not explained in an equally detailed manner. In response, middle school textbook publishers quickly produced curricular materials intended to align with the need for informal inferential reasoning in grade 7. Yet, many teachers, especially at the middle school level, do not have experience teaching informal inference. We argue that guidance is needed on how to assess the fidelity of inferential reasoning tasks contained within these curricular materials. While this may seem to be a narrow focus, inferential reasoning is a key learning goal of statistical education and incorporates knowledge of core statistical concepts and reasoning about data distributions. In this paper, we describe a framework we developed for characterizing informal inferential reasoning tasks based on recommendations of statistics education research, and then share how we analyzed tasks from three widely available seventh grade textbooks.

Informal Inferential Reasoning

In order to define and situate informal inferential reasoning for the purposes of this paper and framework, two broader concepts must be described: statistical inference and statistical reasoning. *Statistical inference* refers to moving beyond the data at hand to make decisions about some wider universe, taking into account that variation is everywhere and conclusions are therefore uncertain (Moore, 2004). *Statistical reasoning* is defined "as the way people reason with statistical ideas and make sense of statistical information" (Garfield & Ben-Zvi, 2004, p. 7). Hence, *inferential reasoning* is the way people make sense of statistical ideas and information with the goal of generating a conclusion that extends beyond the data at hand.

Bartell, T. G., Bieda, K. N., Putnam, R. T., Bradfield, K., & Dominguez, H. (Eds.). (2015). *Proceedings of the 37th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. East Lansing, MI: Michigan State University.

Generally, two types of problems fall under the broad definition of inferential reasoning: (a) generalizing from samples to populations, and (b) comparing samples to determine significant differences in populations (Garfield & Ben-Zvi, 2008). While students can address these problems with formal hypothesis tests, they can also formulate responses based on informal approaches that do not involve set procedures, but rather coordination of prior knowledge, statistical concepts, and the context of the problem. *Informal inferential reasoning* allows students in upper-elementary grades to engage and successfully draw inferences (Stohl & Tarr, 2002; Watson, 2002; Watson & Moritz, 1999).

Informal Inferential Reasoning Task Framework

Building upon the work of Zieffler, Garfield, DelMas and Reading (2008) and others, we developed a framework for assessing informal inferential tasks in middle school mathematics textbooks that both embodies the key recommendations for developing informal inferential reasoning and captures common trimming attributes, which lower the cognitive demand and opportunities to learn (See Table 1). While the recommendations from leaders in statistics education and other disciplines provide a comprehensive list of requirements for inferential reasoning tasks, our framework acknowledges a spectrum within each task dimension (i.e., inference, ill-structured, open-ended, context, and visual representation) that reveals nuances in tasks and ultimately pedagogical choices made by textbook authors and publishers that directly impact students' opportunities to learn

Table 1: Informal Inferential Reasoning Task Framework

Task Dimension	Low (Deterministic) - Limited/No reasoning required	Medium – Some inferential reasoning required	High – Inferential reasoning required
Inference	A population is utilized or the type of the data is unspecified. No requirement is needed to infer beyond data provided.	Sample data is utilized with the acknowledgement of variation.	Sample data is utilized with the acknowledgement of variation, and students are required to infer beyond the data at hand.
Ill-Structured	A prescribed procedure is desired with specified descriptive statistics computations.	A procedure exists that can be adapted in order to coordinate core statistical concepts with a choice of statistical measures.	Coordination of core statistical concepts is required to fully address the task without a prescribed solution path.
Open-Ended	Only one acceptable or "correct" solution exists.	Multiple numerical solutions with similar interpretations are possible or limited numerical solutions exist with a variety of possible interpretations.	Multiple numerical solutions are possible and a variety of conclusions.
Context	The task can be addressed fully by removing the context.	The context is helpful for generating an inference, but not required.	The problem context must be considered in order to generate a viable inference.
Visual Representation	Visual representations are neither provided nor encouraged.	Visual representation are provided or created, but mask the original data.	Raw data is provided and organized in graphical representations.

and reason about statistics. For each task dimension, we created a tiered set of categories based on the level of inferential reasoning required for the task: low (deterministic), medium, and high.

Inference

The first task dimension, inference, relates to how sample and population data are presented and utilized in tasks. Based upon a synthesis of research from educational psychology, science education, and mathematics education, statistics educators recommend informal inferential reasoning tasks require students to:

> 1) make judgments, claims, or predictions about a population based on samples, but not using formal statistical procedures or methods, 2) draw on, utilize, and integrate prior knowledge (formal or informal) to the extent that this knowledge is available, and 3) articulate evidence-based arguments for judgments, claims, or predictions based on samples. (Zieffler et al., 2008, p. 46-47).

A key facet of these recommendations relates to the need for students to experience and think about the differences between complete populations and samples. If a complete population is provided or the source of the data is unknown, then the task does not require inferential reasoning and is reduced to simply computing the differences in measures of center or another statistic of interest to draw a concrete and certain conclusion. Only through sample data is uncertainty introduced, which is the nature of statistics versus a deterministic mathematical problem.

Ill-Structured

Ill-structured tasks require informal reasoning versus applying formal approaches. Reasoning effectively to generate informal inferences requires prior knowledge of core statistical ideas, such as measures of center, variation, skew, outliers, shape of data distribution, and sample size, and an understanding of the relationships between them (Garfield & Ben-Zvi, 2007). Many statistical questions require coordination of both a measure of central tendency, such as mean or median, with a measure of variation such as range, interquartile range, or mean absolute deviation (MAD). In addition, middle school textbooks include tasks that require coordinating and comparing two measures of center, two measures of variation, or other combinations.

The second criterion for this dimension relates to the extent that the task is either well- or ill-defined in nature. Informal approaches to reasoning are needed when problems either do not align with known solution methods or are presented before students possess the knowledge of such methods. One would expect that students possess varying repositories of knowledge, which would result in a diversity of solution strategies when administered similar inferential reasoning tasks. This knowledge might consist of prior statistical knowledge, life experiences related to the context, and informal reasoning skills. As Means and Voss (1996) state, “Informal reasoning assumes importance when information is less accessible, or when the problems are more open-ended, debatable, complex, or ill-structured, and especially when the issue requires that the individual build an argument to support a claim” (p. 140).

When students approach ill-structured problems, they generally progress through four phases: problem structuring, preliminary design, refinement, and detailing (Goel, 1992). As ideas are flushed out in more detail, students become more committed to their solution strategy. The omission of one correct answer or lack of problem constraints is the key factor for encouraging informal reasoning. Watson and Moritz (1999) describe an iterative process that students embarked upon when comparing two data distributions involving: comparing measures of center, then considering other characteristics of the data distribution such as skew or range, and finally coordinating all possible data comparisons together to produce a detailed and integrated response. These steps provide a view into students’ statistical reasoning beyond traditional tasks that are highly structured in nature and

Bartell, T. G., Bieda, K. N., Putnam, R. T., Bradfield, K., & Dominguez, H. (Eds.). (2015). *Proceedings of the 37th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. East Lansing, MI: Michigan State University.

seek a predetermined solution. The ranking for this category requires that no prescribed solution path is provided in advance and that students must compare at least two core statistical concepts.

Open-Ended

Open-ended tasks directly connect to the goal of eliciting informal approaches to inferential tasks (Bakker, 2004; Cobb, McClain, & Gravenmeier, 2003; Garfield & Ben-Zvi, 2007; Watson, 2002; Watson & Moritz, 1999). According to Leathman, Lawrence, and Mewborn (2005), open-ended problems "elicit reasoning, problem solving, and communication" (p. 413). Characteristics of high quality, open-ended tasks include the involvement of significant mathematics, the potential to solicit basic to sophisticated responses, and a balance between too much and too little information. Clearly, the bounds of ill-structured tasks and open-ended tasks overlap to some degree as the descriptions of both include common characteristics.

Many teacher-researchers initially introduce open-ended tasks to hone students' thinking and reasoning about a situation. Through whole class discussion, the open-ended tasks become closed as taken-as-shared meanings develop (e.g. Cobb, 1999). In one study, students were asked to determine which of two ambulance service providers was better and provide justification for their reasoning (Cobb, McClain, & Gravenmeier, 2003). During a lengthy whole class discussion, students determined a process for reasoning about the information provided and agreed upon a final conclusion. Hence, the initially open-ended task became closed through the instructional process of establishing norms for acceptable justification.

By understanding this natural instructional sequence of tasks initially being open-ended in nature and over time becoming close-ended through the course of learning and whole class discussions, we anticipate not all tasks in a textbook would meet this requirement within an instructional unit. As students see relationships between tasks and establish ways of reasoning, the variety of conclusions will decrease with experience. However, if prescribed answers are provided for all inferential tasks, then the textbook is not allowing adequate room for students to engage in informal reasoning. Therefore, open-ended tasks require students to decide what is relevant and what constitutes acceptable justification without prior instruction. For example, if a textbook supports a range of answers as acceptable or incudes a clause, such as "Answer will vary", then the task is deemed to be open-ended in nature. In addition, high quality, open-ended tasks require some level of justification or explanation to accompany the conclusion based on the selected relevant information. Therefore, we attend to both the open-ended nature of the response and the need for justification.

The Role of Context

The authors of the *GAISE* recommendations (Franklin et al., 2007) state, "In mathematics, context obscures structure. In data analysis, context provides meaning" (p. 7). Hence, the use of context is the norm in statistics education and instructors commonly introduce data sets in relation to some real-world phenomena or situation. However, the way statistics educators use context in their tasks varies substantially. On one hand, several have created problem scenarios familiar to students in an effort to increase accessibility and leverage prior knowledge and experiences (Bakker, 2004; Garfield & Ben-Zvi, 2007; Watson & Moritz, 1999; Watson, 2002; Watson, 2008). For example, Watson created a sequence of tasks based on measures of actual students' heart rates and arm-span lengths. Creating data sets close to the knowledge and experiences of students helps focus the tasks on the reasoning process.

On the other hand, some researchers advocate tasks based on real-world contexts. Cobb (1999) and Cobb, McClain and Gravemeier (2003) created a variety of real-world contexts such as ambulance response times, success of speed traps, effectiveness of AIDS treatments, battery life spans, SAT scores based on school expenditure, and response time versus alcohol intake. Cobb, McClain and Gravemeier (2003) state that students must find the context of the problem both plausible and important before they will engage in reasoning about the data. In our framework, we attend to the inclusion of context and the role it plays in terms of generating an inference. Because we cannot be certain of which contexts will be either familiar or engaging to students, we focus only on the role of the context in the problem. If the context can be stripped away and/or ignored, the task is coded as low on the framework. If the context facilitates reasoning about the task, but is not needed to generate a response, then it is coded medium. Tasks that require attending to the context and incorporating it are ranked high.

Visual Representations

Visual representations shift students' thinking away from local attributes or summary statistics towards global characteristics and relationships. Tasks involving small sets of data (n<50) encourage the use of dot plots and bar graphs to depict the data distributions (Bakker, 2004; Garfield & Ben-Zvi, 2007; Watson 2002; 2008; Watson & Moritz, 1999). In addition to shifting students' thinking toward the entire distribution versus individual data values, visual representations facilitate coordination of core statistical concepts in a way that is extremely difficult with only summary statistics and little prior experience with statistical reasoning. The most useful representations for novices are graphical displays that reveal the raw data, in addition to organizing it visually, such as dot plots (Franklin et al., 2007). Therefore, we privilege representations that reveal the raw data and do not restrict the students' reasoning.

In the cases where only raw data is provided without a graphical display or a prompt to create a graphical display, the task is coded low. If the task contains graphical displays that mask the original data values (e.g. box-plots), it is coded medium. We acknowledge that box-plots serve an important role in inferential reasoning, by providing a lens in which to view the data that is useful. However, reasoning is restricted to some degree, as characteristics of the original data distribution are hidden from view. Lastly, if the data values are provided or generated by the students and visual representations are either provided or encouraged, the task is coded high.

Application of the Framework

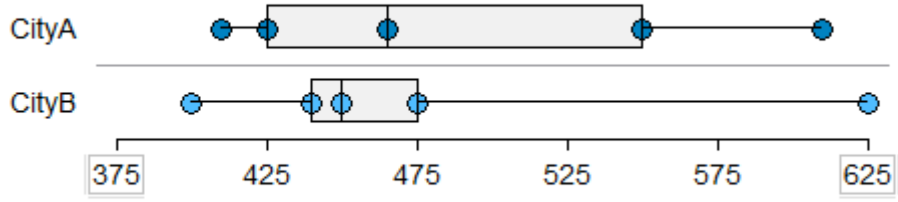
Analysis of Teacher Materials

We examined the teacher's editions of three commonly used 7th grade textbooks and identified the chapter(s) on statistics. In the chapter(s), the textbooks often reference examples for students' problems. Therefore, we analyzed the task based on the cited example. We acknowledge that hypothetically the task could be solved in a variety of ways; however, the example implies a set procedure path. In addition, if the answer key requires only a numerical answer, the task was classified as close-ended. Finally, if the task could be completed fully without considering the context, we coded the task low. The purpose of the following section is not to provide representative or typical tasks of the textbooks, but rather to demonstrate how the framework can be applied to a variety of informal inferential tasks found in CCSS-M aligned grade 7 textbooks.

Applying the framework, the task in Figure 1(problems 1, 2, and 3 inclusive), does not meet the requirement for inference since the source of the data is unspecified. One might assume this representation includes all the data of rental costs for each city, as there is no verbiage to the contrary. In regard to the task being ill-structured, prior examples in the textbook provide an approach to this problem of comparing the inner quartiles and the ranges of the box-plots. Since the inner quartile of CityB is smaller than CityA, yet the range of City B is larger than CityA, students will need to decide how to proceed. Therefore, this task is medium in terms of being ill-structured. A specified path exists but can be modified to accommodate coordination of core statistical concepts based on student's discretion. Next, the task is high in terms of being open- ended in nature, as the textbook notes that answers will vary. Depending on the decisions made when comparing CityA to

Bartell, T. G., Bieda, K. N., Putnam, R. T., Bradfield, K., & Dominguez, H. (Eds.). (2015). *Proceedings of the 37th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. East Lansing, MI: Michigan State University.

To answer the following problems, use the box-and-whisker plots of apartment rentals in two different cities.



1. *Which city has a greater median apartment rental cost?*
2. *Which city has a greater interquartile range of apartment rental costs?*
3. *Which city appears to have a more predictable apartment rental cost?*

Figure 1: Task adapted from Holt McDougal (2012)

CityB, students may arrive at different justifications. The context of the problem does not appear to be needed or facilitate reasoning, so it is rated low. Although a visual representation is provided, the original data is masked, leading to a medium ranking. Overall, we conclude that this task provides some opportunities for students to engage in aspects of informal inferential reasoning, but falls short of requiring all aspects.

The double dot plot below shows the quiz scores out of 20 points for two different class periods. Compare the centers and variations of the two populations. Round to the nearest tenth. Write an inference you can draw about the two populations.

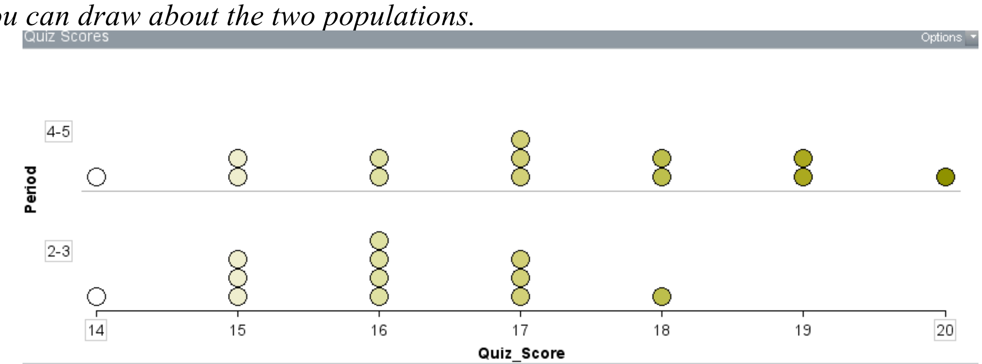


Figure 2: Task adapted from Glencoe (2013)

Applying the framework to this task, we conclude that it does not meet the requirement of an inferential task. The task implies that the dot plots represent the population for the two groups of class periods. In regard to the task being ill-structured, prior examples in the textbook provide a procedure of first comparing mean values and then comparing MADs. Students are steered to conclude that periods 4-5 have a higher mean and a larger MAD or more variation. Therefore, periods 4-5 scored higher on average, but the scores varied more and were spread out. In terms of being open-ended, the task is low because one correct answer is noted in the teacher's edition. In addition, the context is not needed for the problem and perhaps inhibits reasoning by grouping the data of two class periods. Lastly, in terms of visual representation, the task ranks high with the raw data visible and organized in a way that facilitates coordination of core concepts and informal reasoning. Overall, this task ranks low in terms of providing students opportunities to informally reason about inference.

Make a Conjecture *The box plots show the distributions of mean weights of 10 samples of 10 football players from each of two leagues, A and B. What can you say about any comparison of the weights of the two populations? Explain.*

Distribution of Means from 10 Random Samples of Size 10
Means

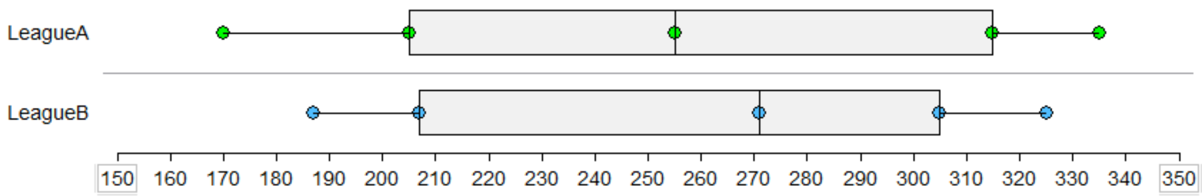


Figure 3: Task adapted from Go Math! (2014)

This task is different from the others as the box-plots are sampling distribution of means, a sophisticated statistical concept that has proven illusive to many tertiary students in introduction to statistics courses. The textbook recommends inferential reasoning with distributions of sample means as a way to reduce variability and make better comparisons, since the means vary less than the original data. Applying the framework to this task, we conclude this task meets the full requirements of an inferential task, as the data are labeled as samples of size 10 and students are asked to generate a conclusion that extends beyond the data at hand. In regard to the task being ill-structured, prior examples in the textbook provide an approach to the problem of comparing the centers of the distributions and looking at the overlapping portions of the inner quartile. Students may or may not understand why this approach works, but it is specified. Hence, we would rank this as low in terms of being ill-structured. Students will note that League B has a higher mean, but the overlapping inner quartiles create ambiguity in terms of which league has higher weight in general. Therefore, the task is close-ended with one correct answer. In addition, the context is not needed for generating the inference. In terms of visual representation, the task ranks medium with a graphic display and no access to the original data. Overall, we would conclude this task does provide some opportunities for students to engage in aspects of informal inferential reasoning, but falls short of requiring all aspects.

Conclusion

With the advent of many new mathematics textbooks claiming to align with national standards and research-based recommendations, tools are needed to assess the fidelity of tasks posed to students. Further, to study the learning effects of first introducing inference through informal approaches followed by formalization, middle school students require authentic experiences with informal inferential reasoning. Without the development and utilization of frameworks based on prior research and educational experiences, we will never know if students have the opportunities to informally generate inferences that later lead to a robust and connected understanding of formal statistics. Finally, we need to hold textbook publishers accountable for providing students with authentic opportunities to sense-make and reason, as outlined by leaders in statistics education.

References

Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal, 3*(2), 64-83.

Carter, J. A., Day, R., Cuevas, G. J., & Malloy, C. (2013). *Glencoe Math, Course 2.* Columbus, OH: The McGraw-Hill Companies, Inc. Glencoe (2013).

Cobb, P. (1999). Individual and collective mathematical development: The case of statistical data analysis. *Mathematical Thinking and Learning, 1*(1), 5-43.

Cobb, P., McClain, K., & Gravemeier, K. P. E. (2003). Learning about statistical covariation. *Cognition and Instruction, 21*(1), 1-78.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report.* Alexandria, VA: American Statistical Association.

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

Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review, 75*(3), 372–396.

Garfield, J. & Ben-Zvi, D. (2004). Statistical literacy, reasoning and thinking: Goals, definitions, and challenges. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 3-15). Dordredcht, The Netherlands: Kluwer.

Goel, V. (1992). A Comparison of Well-structured and Ill-structured Task Environments and Problem Spaces. *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Holt McDougal (2012). *Holt McDougal mathematics grade 7: Common core edition.* Houghton Mifflin Harcourt.

Burger, E. B., Dixon, J. K., Kanold, T. D., Larson, M. R., Leinwand, S. J., & Sandoval-Martinez, M. E. (2014). *Go Math! Teacher edition grade 7.* Orlando, FL. Houghton Mifflin Harcourt.

Konold, C., & Pollatsek, A. (2002). Data analysis as a search for signals in noisy processes. *Journal for Research in Mathematics Education, 24*, 392-414.

Leathman, K. R., Lawrence, K., & Mewborn, D. S. (2005). Getting started with open-ended assessment. *Teaching Children Mathematics, 11*(8), 413-419.

Means, M. L., & Voss, J. F. (1996). Who reasons well? Two studies of informal reasoning among children of different grade, ability, and knowledge levels. *Cognition and Instruction, 14*(2), 139-178.

Moore, D. (2004). *The basic practice of statistics* (3rd ed.). New York: W. H. Freeman.

National Governors Association Center for Best Practices and Council of Chief State School Officers. (2010). *Common core state standards: Mathematics standards.* Retrieved January 17, 2013 from <http://www.corestandards.org>.

Stohl, H., & Tarr, J.E. (2002). Developing notions of inference with probability simulation tools. *Journal of Mathematical Behavior*, *21*(3), 319-337.

Watson, J. M. (2008). Exploring beginning inference with novice grade 7 students. *Statistics Education Research Journal, 7*(2), 59-82.

Watson, Jane M. (2002). Inferential reasoning and the influence of cognitive conflict. *Educational Studies in Mathematics, 51,* 225 – 256.

Watson, J. M., & Moritz, J. B. (1999). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics, 37*, 145-168.

Wild, C., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67(3),* 223-265.

Zieffler, A., delMas, R., Garfield, J., & Gould, R. (2007, August). *Studying the development of college students' reasoning about statistical inference.* Paper presented at the Fifth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-5), University of Warwick, UK.