MIDDLE AND SECONDARY TEACHERS' INFORMAL INFERENTIAL REASONING

Christopher Engledowl
New Mexico State University
chriseng@nmsu.edu

James E. Tarr
University of Missouri
TarrJ@missouri.edu

This study examined middle and secondary mathematics teachers' knowledge structures and informal inferential reasoning (IIR). Using task-based clinical interviews (Goldin, 1997) and cross-case analysis, nine teachers responded to four LOCUS assessment tasks (Jacobbe, 2016). Responses were used to construct knowledge structure maps for measures of center, spread, and shape (Groth & Bergner, 2013). Teachers' IIR was analyzed for the appropriateness of responses (Means & Voss, 1996) and key components of IIR were identified. Teachers with more connected knowledge structures and fewer undesirable knowledge elements exhibited more acceptable forms of IIR. Although teachers engaged in the inference and data components of IIR (Makar & Rubin, 2009), they rarely referenced uncertainty. Implications for teacher education and future research are discussed.

Keywords: Data Analysis & Statistics, Teacher Knowledge, Cognition

Globally, there has been an increased recognition of the need for statistical literacy. This is particularly evident in the U.S. from the increased number of statistics standards included in the Common Core State Standards for Mathematics (CCSSM) and from other standards reform documents (e.g., Franklin et al., 2007). This push is due, in part, to consistent findings that tertiary students struggle to improve in their statistical reasoning abilities throughout introductory courses (e.g., delMas, Garfield, Ooms, & Chance, 2007). Not surprising, the majority of K-12 mathematics teachers in the U.S. feel unprepared to teach statistics content, despite completing at least one college level statistics course (Banilower et al., 2013) and despite the consistent calls for mathematics teacher education reform over the past twenty years (Conference Board of the Mathematical Sciences, 2001, 2012; Franklin et al., 2015). More importantly, standards reform documents, such as the CCSSM, require teachers to engage their students in informal ways of reasoning, which is not typically found in tertiary courses (Garfield, DelMas, & Zieffler, 2012), thereby leaving many teachers unable to draw on experience to support such teaching. Several studies have found great success in engaging students in informal inferential reasoning (IIR) as a way to promote students' statistical literacy across middle and secondary grades (see for example, Makar & Ben-Zvi, 2011, special issue in *Mathematical Thinking and Learning*; Pratt & Ainley, 2008, special issue in *Statistics Education Research Journal*). However, there has not been an associated level of research into teachers' engagement in IIR, leading to recent claims that research connecting teachers' content knowledge and IIR is a "critical area for future research" (Langrall, Makar, Nilsson, & Shaughnessy, 2017, p. 517). This study examined teachers' knowledge structures for measures of center, spread, and shape as it related to their engagement in tasks designed to elicit IIR. The goal was to understand the ways in which teachers' knowledge structures may be constructed, and how they support the ways teachers engage in IIR, in the context of task-based clinical interviews. This work adds to research base on teachers' IIR—a relatively thin area compared to research on students.

Theoretical Framework

This study draws on the work of Zieffler and colleagues (2008), who defined informal inferential reasoning (IIR) as "the way in which students use their statistical knowledge to make

Hodges, T.E., Roy, G. J., & Tyminski, A. M. (Eds.). (2018). *Proceedings of the 40th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Greenville, SC: University of South Carolina & Clemson University.

arguments to support inferences about unknown populations based on observed samples" (p. 44). Their framework for IIR describes three essential components—making claims about populations from samples, utilizing prior knowledge, and using evidence-based arguments to support claims about populations (Zieffler et al., 2008, p. 45). We further expanded this framework, drawing upon Rossman (2008), to include inferences made about causality between variables in addition to claims about populations from samples.

In considering the role of knowledge in this IIR framework, we took the stance of Franklin and colleagues (2007) who assumed that knowledge and informal reasoning develop alongside one another. Furthermore, in Zieffler and colleagues' literature review, they found that informal reasoning did not improve with "maturation, education, or life experience" (2008, p. 44). Therefore, we theorized that IIR occurs at the intersection of statistical knowledge and informal reasoning, and neither knowledge nor reasoning are required for the development of the other.

Research Questions

This study addressed the following two research questions: What knowledge structures do middle level and secondary mathematics teachers have regarding center, spread, and shape of distributions (RQ1)? How do teachers' knowledge structures support IIR (RQ2)?

Method

Setting and Participants

A stratified purposeful sample (Patton, 2002) of nine practicing middle and secondary mathematics teachers each participated in two task-based clinical interviews. Teachers were required to have taught statistics content that included data analysis explicitly using measures of center, spread, and shape of distributions. Teachers were also chosen in order to obtain four strata: 1) statistics taught as a unit within middle level mathematics ($N = 3$), 2) statistics taught as a unit within secondary mathematics ($N = 2$), 3) Non-Advanced Placement (AP) Statistics ($N = 2$), and AP Statistics ($N = 2$). There was wide variability in teachers' backgrounds that was not concentrated in any one strata, all teachers had Master's degrees in education-related fields, and all had completed at least one tertiary course in statistics.

Data and Analysis

Task-based clinical interviews were conducted with each teacher, using two tasks for each of two 60–90 minute video recorded interviews. The four tasks were selected from released items from the *Levels of Conceptual Understanding of Statistics* (LOCUS) assessment (Jacobbe, 2016) in order to align with Huey and Jackson's IIR task framework (2015) to maximize the potential for observing IIR engagement. Due to space limitations, the tasks are not included, but they were: New Year's Day Race, Tomatoes and Fertilizer, Extended School Day, and Jumping Distances. Teachers' written and verbal responses aided in answering both research questions. Interview protocols were developed, following the suggestions of Ginsburg (1981) that questions should require reflection, determine the seriousness of responses, confirm that participants understood the question, and evaluate the strength of belief of responses by challenging them. Moreover, we followed suggestions from Goldin (1997) that follow-up questions be non-directive and that the protocol anticipate as many contingencies as possible. Contingencies were more completely anticipated by using LOCUS tasks because a range of student responses are included along with each released task. Due to space limitations, instruments and tasks will not be shown here but will be provided during the presentation.

Data analysis first involved transcribing each interview and coding teachers' responses for constructed knowledge elements related to center, spread, and shape of distributions (RQ1).

Connections between knowledge elements were then also hypothesized through open-coding across all tasks. Node-link diagrams were used in order to visually represent each teacher's knowledge structure. We followed a similar method to that of Groth and Bergner (2013) that involved mapping Partially-Correct Constructions for each participant. Thus, teachers' responses were further classified as either *desirable* or *undesirable*. A *desirable* constructed knowledge element is one that has been identified in the literature as having the potential for supporting the development of other knowledge elements or more complex depictions of that knowledge element. An *undesirable* element is one that may support the development of inconsistent or disconnected knowledge elements, or not allow for a *desirable* element to be constructed from it. *Desirable* elements were visually represented by blue rectangles, and *undesirable* elements with yellow rectangles with rounded vertices. These value laden terms should not be misconstrued as a teacher's lack of expertise because an *undesirable* element does not necessarily imply that it is inherently incorrect. Observed connections between knowledge elements were depicted with a double-headed arrow. After all knowledge maps had been constructed, a within-case analysis was carried out to confirm each knowledge structure, and then a cross-case analysis was carried out to identify common themes across structures to categorize types of structures.

To address RQ2, we drew upon the work of Means and Voss (1996) to first identify teachers' arguments—their claims and reasons for them—and if they were supported by *acceptable* or *unacceptable* evidence. Next, we employed Makar and Rubin's informal statistical inference framework (2009) to identify which of the three components, evidence of IIR, were observed in teachers' arguments—generalization beyond the data, data as evidence, and probabilistic language. A cross-case analysis (Creswell, 2013) was carried out to identify categories of types of IIR. Lastly, types of IIR were compared with types of knowledge structures to characterize how teachers' knowledge structures may support their IIR.

Findings

Knowledge Structures for Center, Spread and Shape

After a cross-case analysis, 3 basic types of knowledge structures were identified (see Figure 1): *desirable-connected* ($N = 3$), *undesirable-connected* ($N = 4$), and *undesirable-disconnected* ($N = 2$). *Desirable-connected* structures included almost no *undesirable* knowledge elements and knowledge elements were highly connected, with connections observed both *within* and *between* knowledge types (center, spread, shape), as can be seen in the case of Rosalynn's knowledge structure as a case of a *desirable-connected* structure in Figure 1. This knowledge structure category was highly consistent across teachers—salient features were observed across all cases. On the other end of the spectrum, *undesirable-disconnected* knowledge structures contained multiple *undesirable* knowledge elements that were observed to be largely disconnected, with no connections observed between center and shape knowledge types. These characteristics were consistent across the two cases of *undesirable-disconnected* structures, as can be seen in the case of Amalia's knowledge structure in Figure 1. Knowledge structures described as *undesirable-connected* also contained multiple *undesirable* knowledge elements, but knowledge elements were highly connected—thus integrating *undesirable* knowledge elements into the overall structure, as can be seen in Ellie's knowledge structure in Figure 1. Knowledge structures in this category were more varied than the other types. For instance, not all structures included direct connections *between* all knowledge types (center, spread, shape), but all included direct connections *between* at least two of the three types. Moreover, the majority of knowledge elements were connected to at least one other knowledge element—reflecting the highly integrated nature of teachers' knowledge. It is of note that in all cases of both *undesirable-*

Hodges, T.E., Roy, G. J., & Tyminski, A. M. (Eds.). (2018). *Proceedings of the 40th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Greenville, SC: University of South Carolina & Clemson University.

connected and *undesirable-disconnected* structures, at least one direct path connecting two elements of center, spread, and shape included at least one *undesirable* knowledge element.

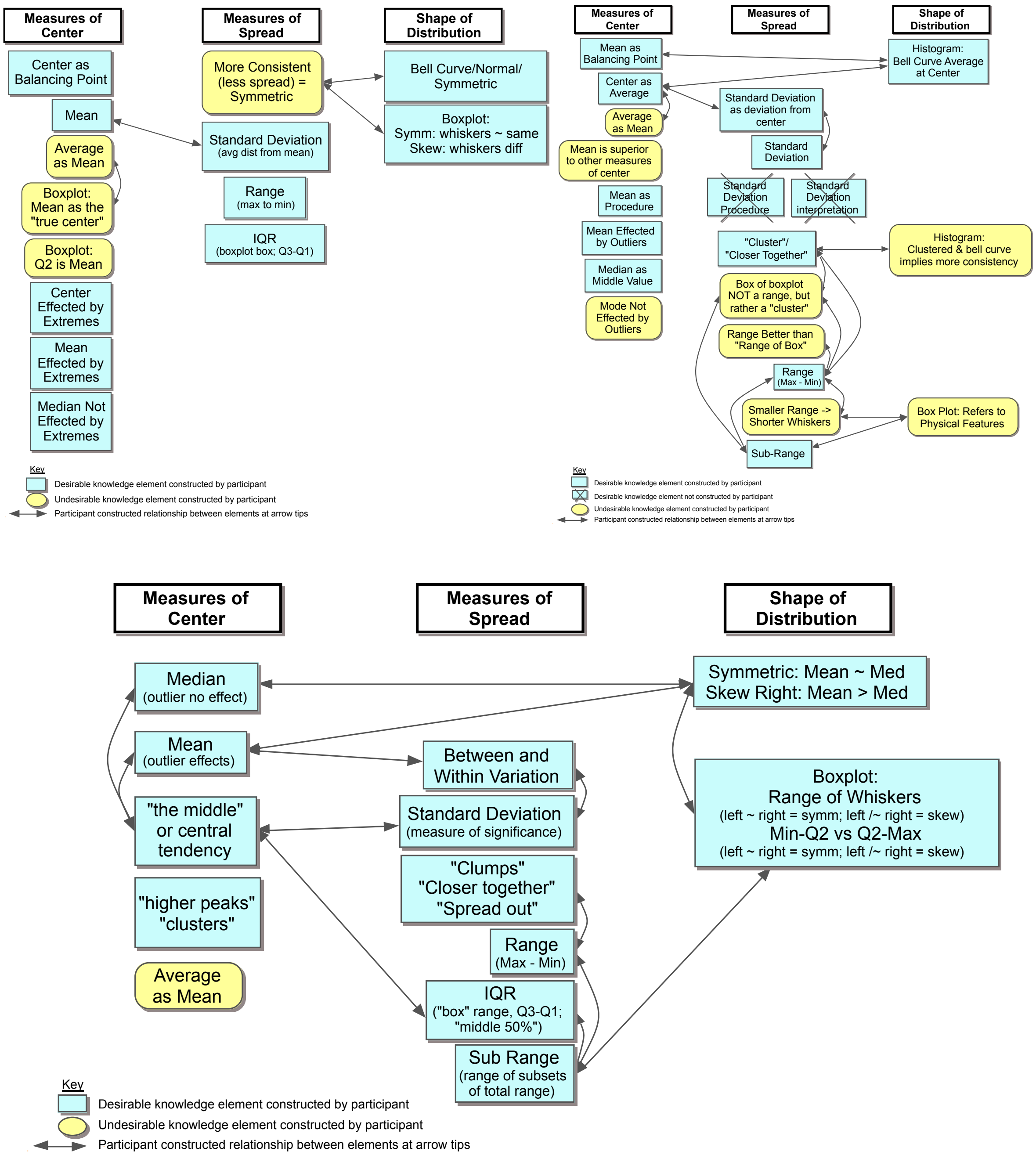

Figure 1. Representative cases of the three types of knowledge structures

Knowledge Structures and Support for IIR

Tasks that did not require engagement in IIR explicitly (e.g., asking only about characteristics of samples) produced different findings than those that did. In particular, teachers more frequently exhibited *acceptable* supports in their arguments when reasoning in these non-

Hodges, T.E., Roy, G. J., & Tyminski, A. M. (Eds.). (2018). *Proceedings of the 40th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Greenville, SC: University of South Carolina & Clemson University.

IIR contexts than in those tasks that demanded IIR (see Table 1). For example, five teachers were observed to reason using *acceptable* supports (i.e., drawing on *desirable* knowledge elements) for at least 75% of their responses (note that teachers were encouraged to offer multiple responses per task per the interview protocol) and two of those (Rosalynn and Tim) reasoned in this way for 100% of their responses. In contrast, within IIR contexts, none of the teachers were observed to reason using *acceptable* supports for at least 75% of their responses. Although teachers' reasoning in IIR contexts largely made use of *unacceptable* supports in their arguments, when the task only required drawing upon center, spread or shape, 8 of the 9 teachers were able to offer at least one response coded as *acceptable*. However, 5 of these 8 teachers simultaneously offered responses coded as *unacceptable*, indicating wider variation in responses for IIR contexts than non-IIR contexts.

Table 1: Comparing Reasoning Types and Knowledge Structures

Non-IIR Reasoning	Teacher	Knowledge Structure	IIR Reasoning
Mostly Unacceptable	Kathy	Undesirable-Disconnected	Mostly Unacceptable IIR
Mixed	Amalia	Undesirable-Disconnected	Mostly Unacceptable IIR
Mixed	Ellie	Undesirable-Connected	Mostly Unacceptable IIR
Mixed	Ruby	Undesirable-Connected	Mostly Unacceptable IIR
Mostly Acceptable	Michaela	Undesirable-Connected	More Unacceptable IIR
Mostly Acceptable	Harrison	Undesirable-Connected	More Unacceptable IIR
Mostly Acceptable	Mike	Desirable-Connected	More Unacceptable IIR
Mostly Acceptable	Rosalynn	Desirable-Connected	Mostly Acceptable IIR
Mostly Acceptable	Tim	Desirable-Connected	Mostly Acceptable IIR

Tasks that required IIR were also coded according to the three components of informal statistical inference described by Makar and Rubin (2009). Two teachers utilized all three components (generalization/causation, data as evidence, probabilistic language) for one response each and 8 of the 9 teachers incorporated both a generalization and data as evidence when their reasoning was coded as *acceptable*. The use of probabilistic language was exceedingly rare, observed in only the two cases where all three components were included, indicating that teachers largely did not consider the deterministic nature of their inferential statements.

Comparing reasoning types to knowledge types, the greatest observable pattern is that teachers with mostly *desirable* elements that were highly integrated also reasoned in *acceptable* forms for the majority of their responses, while teachers with many *undesirable* elements that were largely *disconnected* tended to reason in *unacceptable* forms for the majority of their responses. A second salient pattern is that those who reasoned in more *acceptable* forms across both contexts (non-IIR and IIR) tended to have more connected knowledge structures. Moreover, the relationship between reasoning types and knowledge structures appeared most evident at opposite ends of the spectrum. To illustrate this, on the Jumping Distances task, participants were asked to compare center, spread and shape of the distances a sample of students jumped by examining a pair of boxplots—one boxplot representing jumping distances for a subset who was provided with a target to jump towards, the other not having a target to jump towards (see Figure 2). The following responses come from Amalia and Rosalynn as they respond to this item:

> *Amalia*: The no target looks to be more symmetric. This one [target plot] is close to be symmetric, but it is slightly skewed. […] The whiskers make me think [the no target plot is] not skewed because they look almost the same length. Whereas this one [target plot] is a little longer on the left [gesturing to both whiskers of target plot] […] But yeah, I was just looking

Hodges, T.E., Roy, G. J., & Tyminski, A. M. (Eds.). (2018). *Proceedings of the 40th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Greenville, SC: University of South Carolina & Clemson University.

at the whiskers, and just the fact that the target group, there is a bigger difference between the largest and the shortest jump, versus here [no target], there's not as much of a difference. (lines 10 and 21)

Rosalynn: The shape of the distribution […] Since you do have a longer whisker on the left, you're looking at a slight left skew for the target group versus a more symmetric distribution for the no target group. […] The left part [gesturing to distance from minimum to median] is more spread out for this one [target] than the right part. (lines 6 and 9)

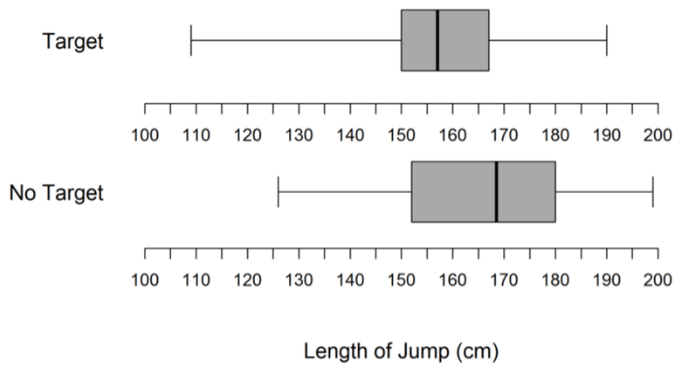

Figure 2. Boxplots provided on Jumping Distances task.

Within these excerpts, you can see that Amalia first begins to describe the distribution's shape using *desirable* knowledge elements—referring to the respective lengths of the whiskers for each plot to hypothesize about the shape. However, she then creates a second argument that draws on an *undesirable* knowledge element—that the smaller range of the no target plot implies more symmetry. Thus, without making contradictory claims, she simultaneously argued in ways coded as both *acceptable* and *unacceptable*. This perspective of spread as being connected to shape in an *undesirable* way led to her *unacceptable* form of reasoning. Amalia's knowledge structure was categorized *undesirable-disconnected* (see Figure 1), and her reasoning was impacted by the integration of *undesirable* elements, thus resulting in more responses coded as *unacceptable*. In contrast, Rosalynn's responses remained in comparison to whether things appeared more evenly spread to the left and right of center, drawing on *desirable* knowledge elements for each response. Therefore, Rosalynn's knowledge structure, categorized as *desirable-connected* (see Figure 1), supported her reasoning in ways that led to more *acceptable* forms of reasoning. Teachers whose knowledge was more integrated yet contained multiple *undesirable* elements, were observed offering both *acceptable* and *unacceptable* forms of reasoning simultaneously—leading to the characterization of having a more mixed form of reasoning.

Discussion

A major finding of this study is that middle and secondary teachers' knowledge structures for center, spread and shape of distributions fell into three categories—*desirable-connected*, *undesirable-connected*, and *undesirable-disconnected*. Looking across these categories, 7 of the 9 teachers were found to have highly interconnected knowledge structures that all made connections among the three knowledge types (center, spread, shape), and some made connections across all three. Moreover, despite many teachers having *undesirable* elements integrated into their knowledge structures, most were observed to have at least one connection between *desirable* elements of center and spread. Although there is evidence that teachers' knowledge structures were highly interconnected, teachers struggled to integrate their knowledge of center, spread and shape. For instance, teachers in this study were largely observed to recognize that connections existed between knowledge element types (less so between center and shape), yet they did not utilize these connections and draw upon multiple knowledge types simultaneously, as prior research has found for students and pre-service teachers (Doerr & Jacob, 2011; Groth & Bergner, 2006; Noll & Shaughnessy, 2012). Moreover, these studies also found that hierarchically higher levels of reasoning required the integration of multiple knowledge elements simultaneously—such as describing a measure of spread in relation to a measure of center—thus, supporting the notion that *desirable-connected* knowledge structures are associated with more *acceptable* forms of reasoning.

Although the theoretical framework for this study does not assume that there is any particular prior knowledge necessary for engaging in IIR—and that knowledge and informal reasoning work together during engagement in IIR—the teachers who engaged in IIR in *mostly acceptable* forms were those with *desirable-connected* knowledge structures. This finding aligns with Makar and colleagues' (2011) claim that statistical knowledge is an important support for IIR. Moreover, as found by Huey (2011) and Watson (2003), teachers in this study tended to focus only on measures of center unless explicitly prompted otherwise.

These results promote the need for statistics teacher education to include explicit opportunities, embedded within tasks, for teachers and preservice teachers to grapple with connections both *within* knowledge types and *between* them. For instance, some teachers in this study believed that outliers were either *always* or *never* ignorable, that the term average *always* implied the arithmetic mean, or that the mean was *always* or *never* a better measure than the median. Situations that confront teachers with these dichotomies will allow them to consider a more nuanced understanding and more flexible knowledge structure. Moreover, teachers in this study largely drew on ideas of center to support inferences, even when tasks did not restrict them in this way. Consequently, tasks designed to engage teachers in IIR should also encourage attention to all of center, spread, and shape—not isolating one type.

More research is needed in these areas in order to explore the kinds of task features that are more likely to encourage integration of all three knowledge types, thus supporting teachers in strengthening their knowledge structures and reasoning. One possible fruitful area of research is to provoke teachers into weighing the evidence of their inferential statements through requiring the construction of multiple inferential statements and responding to multiple inferential statements to explore their validity. This will allow engagement in the statistics practice described in the *Statistical Education of Teachers* report as being able "to compare the plausibility of alternative conclusions and distinguish correct statistical reasoning from that which is flawed" (Franklin et al., 2015, p. 14).

A final needed area of research is in explicitly engaging teachers in, and drawing attention to, the use of probabilistic language. A profound lack of attention to this component of IIR was observed in this study. Unlike mathematics, where deterministic statements are the norm, inferences in statistics are predictions and therefore come with a degree of uncertainty.

References

Banilower, E. R., Smith, P. S., Weiss, I. R., Malzahn, K. A., Campbell, K. M., & Weis, A. M. (2013). *Report of the 2012 national survey of science and mathematics education*. Chapel Hill, NC: Horizon Research, Inc.

Conference Board of the Mathematical Sciences. (2001). *The mathematical education of teachers*. Washington, DC: American Mathematical Society and Mathematical Association of America.

Conference Board of the Mathematical Sciences. (2012). *The mathematical education of teachers II*. Providence, RI and Washington, DC: American Mathematical Society and Mathematical Association of America.

Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches* (3rd ed.). SAGE.
delMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal*, *6*(2), 28–58.
Doerr, H. M., & Jacob, B. (2011). Investigating secondary teachers' statistical understandings. In M. Pytlak, T. Rowland, & E. Swoboda (Eds.), *Proceedings of the Seventh Congress of the European Society for Research in Mathematics Education* (pp. 776–786). University of Rzeszów, Poland.
Franklin, C., Bargagliotti, A. E., Case, C. A., Kader, G., Scheaffer, R., & Spangler, D. A. (2015). *The Statistical Education of Teachers*. Retrieved from http://www.amstat.org/education/SET/SET.pdf
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*. *American Statistical Association*. Alexandria, VA. Retrieved from http://www.amstat.org/Education/gaise/GAISEPreK-12_Full.pdf
Garfield, J., DelMas, R., & Zieffler, A. (2012). Developing statistical modelers and thinkers in an introductory, tertiary-level statistics course. *ZDM - The International Journal on Mathematics Education*, *44*(7), 883–898.
Ginsburg, H. (1981). The clinical interview in psychological research on mathematical thinking: Aims, rationales, techniques. *For the Learning of Mathematics, 4*(3), 1–11.
Goldin, G. A. (1997). Chapter 4: Observing mathematical problem solving through task-based interviews. *Journal for Research in Mathematics Education. Monograph*, *9*(Qualitative Research Methods in Mathematics Education), 40-62.
Groth, R. E., & Bergner, J. A. (2006). Preservice elementary teachers' conceptual and procedural knowledge of mean, median, and mode. *Mathematical Thinking and Learning*, *8*(1), 37–63.
Groth, R. E., & Bergner, J. A. (2013). Mapping the structure of knowledge for teaching nominal categorical data analysis. *Educational Studies in Mathematics*, *83*(2), 247–265. https://doi.org/10.1007/s10649-012-9452-4
Huey, M. E. (2011). *Characterizing middle and secondary preservice teachers' change in inferential reasoning*. University of Missouri, Columbia, MO.
Huey, M. E., & Jackson, C. D. (2015). A framework for analyzing informal inferential reasoning tasks in middle school textbooks. In T. G. Bartell, K. N. Bieda, R. T. Putnam, K. Bradfield, & H. Dominguez (Eds.), *Proceedings of the 37th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 431–438). East Lansing, MI: Michigan State University.
Jacobbe, T. (2016). Levels of Conceptual Understanding in Statistics (LOCUS). Retrieved April 18, 2016, from https://locus.statisticseducation.org/
Langrall, C. W., Makar, K., Nilsson, P., & Shaughnessy, J. M. (2017). The teaching and learning of probability and statistics: An integrated perspective. In J. Cai (Ed.), *Compendium for Research in Mathematics Education* (pp. 490–525). Reston, VA: National Council of Teachers of Mathematics, Inc.
Makar, K., Bakker, A., & Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning*, *13*(1–2), 152–173.
Makar, K., & Ben-Zvi, D. (2011). The role of context in developing reasoning about informal statistical inference. *Mathematical Thinking and Learning*, *13*(1–2), 1–4.
Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, *8*(1), 82–105.
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.
Noll, J., & Shaughnessy, J. M. (2012). Aspects of students' reasoning about variation in empirical sampling distributions. *Journal for Research in Mathematics Education*, *43*(5), 509–556.
Patton, M. Q. (2002). *Qualitative Research and Evaluation Methods* (3rd ed.). Thousand Oaks, California: Sage Publications Inc.
Pratt, D., & Ainley, J. (2008). Introducing the special issue on informal inferential reasoning. *Statistics Education Research Journal*, *7*(2), 3–4.
Rossman, A. (2008). Reasoning about informal statistical inference: One statistician's view. *Statistics Education Research Journal*, *7*(2), 5–19.
Watson, J., & Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal*, *2*(2), 3–46.
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.

Hodges, T.E., Roy, G. J., & Tyminski, A. M. (Eds.). (2018). *Proceedings of the 40th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Greenville, SC: University of South Carolina & Clemson University.