Evaluating Statistical Reasoning of College Students in the Social and Health Sciences with Cognitive Diagnostic Assessment

Ying Cui

Mary Roduta Roberts

Andrea Gotzmann

Centre for Research in Applied Measurement and Evaluation (CRAME),

University of Alberta

Evaluating Statistical Reasoning of College Students in the Social and Health Sciences with Cognitive Diagnostic Assessment

Given the importance of statistics to today's society and the continued need for statistics as the information age evolves, the acquisition of statistical knowledge and skills have become more prominent in undergraduate and graduate programs at the tertiary level of education. Equipped with different levels of mathematical skills, many students in the social and health sciences find statistical courses to be challenging and consider them among the most difficult courses in their programs of study. Students, particularly those with limited mathematical backgrounds, tend to show high levels of anxiety when confronted with statistical ideas and activities in both instructional and evaluative situations, which in turn impact negatively on their learning and performance in statistics (Elmore, Lewis, & Bay, 1993; Lalonde & Gardner, 1993; Onwuegbuzie & Seaman, 1995; Zeidner, 1991).

Social and health science students often experience difficulties in understanding important statistical concepts and in applying these concepts in problem-solving situations. Research in statistics education indicates that whereas some students might be able to mechanically follow the computational steps of statistical procedures and be able to plug quantities into formulas, they often fail to fully understand the theoretical rationales underlying the computations (e.g., Jolliffe, 1991). As a result, these students cannot apply what they have memorized or "learned" to novel problems and what they have learned tends to fade quickly once the course is finished.

Over the past decade or so there has been an increasingly strong call for shifting the focus of statistics education from formulas, procedures, and computation skills to statistical literacy, reasoning, and thinking (Ben-Zvi & Garfield, 2004). The desired outcome is to help students develop and sustain the ability to interpret and critically evaluate statistical information presented in the literature and provide students with the ability to reason and think in a probabilistic and statistical manner. One of the recommendations endorsed by the Board of Directors of the American Statistical Association for teaching introductory college statistics course is to use assessments to improve and evaluate learning (Franklin & Garfield, 2006). To achieve this, statistics educators must improve the quality of assessments they use to evaluate student learning and provide timely and informative feedback so students can use this information to improve their learning. It is critical for the instructors of beginning statistics course for social and health science students to have a deep understanding of the nature and development of statistical thinking and reasoning, the types of difficulties and misconceptions that students tend to have, and what to look for in the assessments of student learning.

In recent decades, cognitive science has made tremendous progress in helping shape perspectives on how people learn and develop expertise. Research in cognitive psychology has indicated that learning is not only associated with an accumulation of knowledge but is also critically dependent on how well the acquired knowledge is structured (Chi, Glaser, & Farr, 1988; Ericsson, 2006; Glaser, Lesgold, & Lajoie, 1987; Leighton & Sternberg, 2003; Mislevy, 2006; Sternberg & Pretz, 2005). The critical difference between novices and experts lie in the tendency of experts to organize their knowledge into schemas that facilitate pattern recognition and rapid retrieval and application of knowledge. To estimate student competency in a particular test domain, therefore, we need more than just to evaluate whether students have gained a sufficient amount of knowledge within the domain. Another important aspect of assessments that must be considered is to evaluate whether students can structure and process knowledge in a way that can be retrieved efficiently when confronted with a problem.

The purpose of this paper is to discuss the usefulness of new cognitive diagnostic assessments in helping evaluate and improve students' statistical reasoning in social and health sciences. The most important feature of cognitive diagnostic assessments is to use a cognitive theory or model to guide the development of test items and to interpret student performance on tests. A cognitive model helps illustrate the way knowledge is represented, structured, and retrieved in the human mind. When the cognitive model is used to guide the assessment design and analysis, inferences about how well students structure and process knowledge can be made. In addition, cognitive diagnostic assessment are aimed at providing specific information about students' cognitive strengths and weaknesses which may, in turn, help teachers make instructional decisions intended to improve student learning.

The paper is divided into four sections. In the first section, we define the term statistical reasoning and differentiate it from closely related terms of statistical literacy and statistical thinking. In the second section, we briefly review cognitive studies of statistical reasoning and focus on our discussion on the current prevalent developmental model that researchers in the field have used to describe an underlying developmental learning trajectory of statistical reasoning in increasing stages of sophistication. In the third section, we discuss how cognitive diagnostic assessments have the potential to lead to an enhanced understanding of how to characterize, evaluate, and improve student learning in statistical reasoning. In the fourth section, we comment on the limitations of the literature for characterizing students' statistical reasoning processes during a problem solving task, and outline future research that is needed to better facilitate the development of cognitive diagnostic assessments in measuring and improving student statistical reasoning.

Definition of Statistical Reasoning

Statistical reasoning is a fairly new research area. As with any new area of research, consensus on the definition and conceptualization of key constructs is an important first step towards creating a stable research base. Ben-Zvi and Garfield (2004) define *statistical reasoning* as “the way people reason with statistical ideas and make sense of statistical information. This involves making interpretations based on sets of data, representations of data, or statistical summaries of data. Statistical reasoning may involve connecting one concept to another (e.g. center and spread), or it may combine ideas about data and chance. Reasoning means understanding and being able to explain statistical processes and being able to fully interpret statistical results.” (p. 7). Garfield (2003) lists six reasoning goals for students: 1) reasoning about data, 2) reasoning about representations of data, 3) reasoning about statistical measures, 4) reasoning about uncertainty, 5) reasoning about samples, 6) reasoning about association.

Statistical reasoning is one of three elements in a taxonomy used to classify learning goals or objectives in statistics education. It is closely related to *statistical thinking*, defined as understanding why and how statistical investigations are conducted and the “big ideas” underlying statistical investigations. Fostering statistical thinking is akin to enabling students to develop skills to solve problems as professional statisticians do. *Statistical literacy* is conceptualized as having the understanding and use of the basic language and tools in statistics. This includes knowledge of basic statistical terminology, symbols, and forms of data representation as well as their interpretation.

Garfield and Ben-Zvi (2007) claim that the three elements of statistical literacy, thinking, and reasoning overlap. An example is DelMas’ (2004) contention that statistical reasoning could be viewed as a form of statistical thinking. The distinction between the two concepts lies in the nature of the task. While statistical thinking may involve knowing *when* and *how* to apply statistical procedures, statistical reasoning involves explaining *why* results were obtained and justifying inferences and conclusions. A natural hierarchy emerges among the three concepts where statistical literacy forms the foundation for statistical reasoning and thinking. The definitions provided by Ben-Zvi, Garfield and colleagues (2004) are an attempt to standardize the meanings of these concepts.

A Review of Literature on Statistical Reasoning

Reasoning is a dynamic cognitive process therefore the progression of a student's reasoning during a given task does not tend to be linear, but rather cyclical. Psychological research on human reasoning has informed statistical reasoning more generally, which is then adapted to reasoning in specific areas in statistics. For example, it is stated in psychological research that humans have easier time reasoning in familiar domains. In such cases, the nature of reasoning errors tends to be more mechanical than rational. This finding can be extended to research on statistical reasoning where reasoning errors of statistical novices may be characterized as more rational than mechanical errors.

However, researchers in statistics education contend that psychological research on learning statistics have focused primarily on the content area of probability and on student misconceptions and misunderstandings with no explicit tie to pedagogy and instruction. It is at this point that the two research agendas diverge. Researchers in statistical education focus on designing studies in the context of education to explicitly address issues of statistical reasoning with direct pedagogical and instructional relevance.

Ben-Zvi and Garfield (2004) summarize current research findings on statistical reasoning across domains such as data analysis, measures of centre, distribution, variability, samples, and sampling distribution. The underlying motivation of these studies is to investigate how students begin to reason about these topics and how reasoning develops over time. For example, from an analysis of interview protocols, DelMas, Garfield, and Chance (2001) identified six categories of students' reasoning about sampling distributions. This included: 1) fluency and understanding of concepts and procedures, 2) identification and use of rules for prediction and explanation, 3) presence or absence of contradictory statements, 4) connecting ideas, concepts, and procedures, 5) awareness of inconsistencies or contradictions, and 6) degree of certainty in choices or statements.

Another important line of research on statistical reasoning focuses on identifying misconceptions and common errors that lead to faulty reasoning. A variety of research methods have been used to identify reasoning errors in the literature. A review of relevant research methods is outlined by Chance & Garfield (2001). Garfield (2000, 2003) lists some common identified errors of students in statistical and probability reasoning including: 1) misconceptions about averages, 2) outcome orientation, 3) belief that good samples must represent a high percentage of the population, 4) the law of small numbers, 5) the representativeness misconception, and 6) equiprobability bias. The reader is referred to Garfield (2003) for specific details of these misconceptions.

Jones, Langrall, Mooney, and Thornton (2004) provide a review of *cognitive models of development* in statistical reasoning. A cognitive model of development in student reasoning predicts qualitative differences in reasoning on the task (see Mooney, 2002). Descriptors of what is expected to be observed in terms of student reasoning on particular tasks are identified. A cognitive model of development can serve as an underlying theoretical framework from which to characterize results from studies investigating how statistical reasoning develops. Jones et. al. argue that cognitive models of students' reasoning are specific to particular content domains or tasks. The rationale behind this is that knowledge is not organized in unitary structures that cut across all kinds of tasks and situations but rather organized within specific domains that are defined by particular content or tasks. Therefore, cognitive models of students' reasoning should incorporate key elements and processes of a content domain by which students grow in their understanding. For example, Jones et al. identify three statistical reasoning processes as applied to exploration of univariate and multivariate data: 1) decision making, 2) prediction, and 3) inference and explanation. Within each of these processes, they describe five cognitive levels that characterize the nature of performance, including idiosyncratic, transitional, quantitative, analytical and extended analytical. Each of these levels is derived from a neo-Piagetian view of development currently espoused by Biggs and Collis' Structure of the Observed Learning Outcome (SOLO) taxonomy (1991). More specifically, idiosyncratic is associated with the early developmental ikonic mode; transitional, quantitative, and analytical are associated with the concrete symbolic mode; and extended analytical is associated with the formal mode.

Cognitive Diagnostic Assessment of Statistical Reasoning

In their paper titled "Assessment in Statistics Education: Issues and Challenges", Garfield and Chance (2000) present an extensive discussion on different possible ways of assessing higher levels of statistical reasoning. They argued that the use of alternative assessment approaches, such as projects, authentic tasks, and critiques, can help instructors gain a better understanding of how well their students think and reason with statistical ideas. However, the enrolments of introductory undergraduate statistics courses are typically large. The demands of teaching a large group of students often prevent the instructor from the use of alternative assessments that require extensive analysis of students' work. Therefore, standardized exams seem to be needed for at least two reasons. First, standardized exams offer time efficiency in terms of test administration, scoring, and reporting. Second, by standardizing the tasks to be performed, the conditions under which they are performed, and the criteria used to interpret the results, standardized tests are considered as being more objective.

As pointed out by Garfield and Chance (2000), items on standardized statistical exams tend to place more emphases on testing students' calculation abilities but pay less attention to measuring students' reasoning skills. Given that scores on these tests directly impact their final grades, students, when preparing the exam, feel the pressure to spend more time practising their procedural and calculation skills, which may result in improved test scores on standardized exams. However, the increased scores do not necessarily reflect an actual improvement in student statistical reasoning and understanding. In addition, standardized exams often fail to provide meaningful information to direct student learning. Typically, overall number-correct or percentage scores, perhaps along with several content-based subscores, are reported. These scores do not provide students with explicit information on their cognitive strengths and weaknesses that they can use to guide their next steps in learning.

Given these limitations with traditional standardized statistical exams, we argue that cognitive diagnostic assessments have the potential to better evaluate student higher-level reasoning skills in standardized settings. Over the past two decades, educational measurement has witnessed a steady growth of scholarly interest and activity on cognitive diagnostic assessments. Research efforts in CDA have been fuelled by the increasing demand, from both researchers and educational stakeholders, for more formative information from educational tests (Huff & Goodman, 2007). Cognitive diagnostic assessments (CDA; Leighton & Gierl, 2007a; Mislevy, 2006), integrates cognitive psychology and educational measurement to enhance learning and instruction. A cognitive diagnostic assessment (CDA) is designed to measure a student's knowledge structures and processing skills. Currently, many educational achievement tests report a small number of content-based subscores, but a CDA reports a profile of scores with specific information about a student's cognitive strengths and weaknesses. This cognitive diagnostic feedback has the potential to guide instructors, parents, and students in their teaching and learning processes.

The most significant aspect of cognitive diagnostic assessment lies in its use of a cognitive model or theory to guide the development of test items and the interpretation of student performance. Creating assessments that are grounded in substantive cognitive theory should yield inferences that are more interpretable, meaningful, and valid for the purposes of informing instruction and learning. Currently, development of a CDA begins with articulation of the construct in question by specifying a cognitive model. A cognitive model in educational measurement refers to a "simplified description of human problem solving on standardized educational tasks, which helps to characterize the knowledge and skills students at different levels of learning have acquired and to facilitate the explanation and prediction of students' performance" (Leighton & Gierl, 2007b, p. 6). The resulting cognitive model serves a dual function. It represents the ordered knowledge, skills, and processes students use to solve items associated with measuring the construct. The model also provides a framework for designing diagnostic items for linking examinees' test performance to specific inferences about cognitive skills. The purpose of the assessment, then, is to characterize student performance in terms of the knowledge and skills mastered based on his or her performance on the diagnostic items.

A cognitive model in the domain of statistics is expected to provide a developmental view of the typical progressions from novices to experts in terms of the development of statistics reasoning skills. This information can help illustrate the understanding that learners normally exhibit at different stages of learning, which shed light on how more competent learners differ from less competent learners in terms of their knowledge structures and processes. Examples of the cognitive model of statistical reasoning are the models developed by Jones et al. and Mooney (2002) to specifically characterize statistical reasoning for elementary and middle school students, respectively. A cognitive model of statistical reasoning is especially valuable for assessment design because it has the potential to help teachers (1) identify the most important cognitive aspects of student learning that are essential for competency in the domain, (2) develop or select test items that support the evaluation of student performances on these aspects, and (3) interpret and report test results in terms of students' current cognitive states of performances and what is most important for their next stage of learning.

To develop cognitive models of student reasoning, the educator/researcher must first identify the construct, making explicit the key underlying knowledge, skills, and processes of the construct. The method could be an expert task or analysis of representative tasks in the given domain. Using an underlying theoretical developmental framework, predictions of performance relative to the identified knowledge, skills and processes of the construct can be outlined. The categories created are qualitatively distinct and validated using empirical student data. Therefore, classification into one of these categories can provide valid diagnostic information about student performance that is grounded in a cognitive model developed with explicit connections to the classroom context.

Cognitive Diagnostic Assessment using Ordered Multiple Choice Items

While statistical problems require the use of formulas and computations, items to be included in a statistical test should focus on assessing student statistical reasoning and thinking. With the guidance of a cognitive model, test items can be carefully designed to provide students with opportunities to demonstrate their understanding, and possible misconceptions, of important statistical concepts and their ability to think and reason with statistics. One assessment design option for directly incorporating models of student cognitive development is the use of ordered multiple choice (OMC) item format for diagnostic assessment (Briggs, Alonzo, Schwab, & Wilson, 2006). The central idea of OMC items is to link the possible answer choices to different developmental levels of student understanding, thereby facilitating the diagnostic interpretation of student item responses. OMC items can be used in standardized tests given that they can be scored quickly and reliably. This type of items provides teachers with an innovative assessment tool to diagnose student learning and guide instructions. The new OMC format represents an endeavour to enhance the diagnostic utility of standardized tests by linking a cognitive model of development to the design of test items.

Overview of developing diagnostic assessments using the OMC format

The first step in creating OMC items is to define the domain of content knowledge to be measured on the assessment. Concepts to be included in the domain will depend on a number of factors including the purpose of the assessment and on the intended uses of assessment results. In the Briggs et. al (2006) example, they created developmental cognitive models in the form of construct maps in the domain of science. More specifically, OMC items were developed for domains that were included in national standards documents for fifth and eighth grades and where there was a body of literature from which to construct developmental progressions and common misconceptions in science learning. The number of developmental levels in the continuum can vary, but each level should be qualitatively distinct from the other levels. For example, common misconceptions that characterize Level 1 understanding can be resolved in Level 2 understanding yielding qualitatively different levels of understanding.

Next, multiple choice items are written to measure the concepts within the construct map. Each item will have response choices that correspond to the different levels of thinking defined within the construct map. The construct map and OMC items are subjected to expert review. Any revisions are made at this time to the construct map, the items, or both, ensuring that the construct map accurately reflects a developmental model of student learning and that the items measure the concepts within the construct map. Once revisions are completed, the OMC items are pilot tested with the intended group of students.

Analysis of student response data can provide evidence of whether the developmental model underlying the construct map is supported empirically. Further refinements to the construct map can be made based on the empirical evidence. For classroom purposes, analysis of the pattern of responses of a student can provide insight into the level of understanding because each response option is tied to specific developmental levels within the cognitive model. In this way, assessments developed using OMC item formats can provide diagnostic information that can be used to guide instructional decisions.

The use of the diagnostic tests in introductory statistics courses can reveal the levels of students' statistical understanding and reasoning. The diagnostic tests can provide diagnostic information about student strengths and weaknesses at different points during the course. Students will not only receive feedback about their overall performance on the test, but will also receive summaries of the missed questions and detailed solutions and explanations. These solutions and explanations can serve as a model for students about how to think and reason using statistics. Hopefully, when provided with more useful feedback about their problem solving on tests, students will later experience less anxiety when confronted with statistical problems.

Summary and General Discussion

The purpose of this paper was to discuss the usefulness of new cognitive diagnostic assessments in helping evaluate and improve students' statistical reasoning in social and health sciences. Statistical reasoning is one of three elements, along with statistical literacy and thinking, that statistics educators are aspiring to teach, and foster development within their students. Despite the importance of statistics to today's society, students in social and health sciences tend to see introductory statistics and more advanced statistics as a set of formulas to learn and, when using the computer, a simple menu to select analytic procedures from. This is partly because what students are assessed on in these courses tends to emphasize procedural and calculation skills, which deviates from the goal of statistics courses to help students be able to think and reason statistically.

It is clear that an alignment of assessment to specific learning goals can best support and foster key educational concepts including statistical reasoning. In order to facilitate this alignment to improve teaching and learning effectiveness, instructors need to understand how students develop statistical reasoning and thinking. This emphasizes the need for findings from cognitive studies of statistical reasoning and thinking to guide the design and implementation of statistics assessment for beginning social and health science graduate students (Ben-Zvi & Garfield, 2004; Fennema & Franke, 1992; Cobb et al., 1991; Resnick, 1983). Statistics education researchers have adapted research findings from psychology related to reasoning in statistics and probability where reasoning is found to be a dynamic process that is domain specific and that is prone to errors of different kinds, depending on the students' familiarity with the domain. Psychological research has provided direction as to the misconceptions and misunderstandings encountered in certain content areas in statistics. However, these misconceptions are often identified in contexts that are separated from the classroom and are not explicitly tied to pedagogy and instruction.

Statistics education researchers have used fairly consistent methods to identify frameworks that guide their studies to yield results that can inform instruction and later, evaluate instructional effectiveness on remediating student errors. More specifically, these researchers have used an underlying developmental framework to predict different qualitative levels of performance in statistics tasks. These levels are applied to specific statistical contexts such as variability and measures of center. Student responses are coded relative to these categories and frameworks modified if the data does not support the initial framework. From these methods, common errors and misconceptions have been identified for some areas.

Although general frameworks have been developed to characterize elementary and middle school students' statistical reasoning (Jones et al., 2000; Mooney, 2002; Mooney, Langrall, Hofbauer, & Johnson, 2001), there is a need for the development of a framework that informs students' statistical understanding and reasoning processes at the university level. The framework should provide a coherent picture of the range of statistical reasoning that university students may have developed. This information can serve as a guideline to direct targeted and efficient instruction and assessment of social and health science students. Before the new framework is used to inform instruction and test design, however, it is critical to empirically evaluate the validity of the framework because it directly links to the accuracy of the inferences made about students. Evidence is needed to demonstrate that the framework indeed reflects how students develop statistical reasoning. Methods commonly used in cognitive psychological research such as protocol analysis (Ericsson & Simon, 1993) can be conducted to collect important psychological evidence for validating cognitive models (Leighton, 2004).

With the guidance of the cognitive framework, cognitive diagnostic assessments can be created to specifically measure the levels of students’ statistical understanding and reasoning of an introductory statistics course for social science and health science students. In this way, instructors would have an innovative tool that they could use early in an introductory course to assist students experiencing difficulty, thereby leading to a more positive learning experience for students and a stronger sound knowledge and skill base for their later learning and use. The development of statistical thinking and reasoning can help students in the social and health science apply statistical procedures accurately and with confidence and validly interpret the obtained statistical results. Equipped with statistical thinking and reasoning skills, students will be better able to complete their theses and dissertations and, later, their own research in their places of employment following graduation.

References

Ben-Zvi, D. & Garfield, J. (2004). Statistical literacy, reasoning, and thinking: goals, definitions, and challenges. In D. Ben-Zvi and J. Garfield, *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking*, Eds., pp.3–16. Dordrecht: Kluwer Academic.

Biggs, J. B., & Collis, K. F. (1991). Multimodal learning and quality of intelligent behavior. In H. A. H. Rowe (Ed.), *Intelligence: Reconceptualization and measurement*, pp. 57–66. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc

Briggs, D., Alonzo, A., Schwab, C., & Wilson, M. (2006). Diagnostic assessment with ordered multiple-choice items. *Educational Assessment*, 11 , pp. 33-64.

Chance, B. L. & Garfield, J. B. (2001). New approaches to gathering data on student learning for research in statistics education. *Statistics Education Research Journal, 1,* 38-41. Retrieved April 25, 2009 at http://fehps.une.edu.au/serj.

Chi, M. T. H., Glaser, R., & Farr, M. J. (1988). *The nature of expertise*. Hillsdale, NJ: Lawrence Erlbaum.

Cobb, P., Wood, T., Yackel, E., Nicholls, J., Wheatley, G., Trigatti, B., & Perlwitz, M. (1991). Assessment of a problem-centered second-grade mathematics project. *Journal for Research in Mathematics Education, 22,* 3–29.

DelMas, R. C. (2004). A comparison of mathematical and statistical reasoning. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, literacy* (pp. 79-95). Norwell, MA: Kluwer Academic Publishers.

DelMas, R. C., Garfield, J., & Chance, B. L. (2001). A framework for the development of students' statistical reasoning. *Statistics Education Research Journal, 1,* 44. Retrieved April 25, 2009 at http://fehps.une.edu.au/serj.

Elmore, P. B., Lewis, E. L., & Bay, M. L. G. (1993) Statistics achievement: A function of attitudes and related experience. Paper presented at the annual meeting of the American Educational Research Association, April (ERIC Document Reproduction Service No. 360 324).

Ericsson K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In: Ericsson KA, Charness N, Feltovich P, Hoffman RR, eds. *Cambridge Handbook of Expertise and Expert Performance* (pp. 685–706). Cambridge, UK: Cambridge University Press.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: The MIT Press.

Fennema, E., & Franke, M. L. (1992). Teacher's knowledge and its impact. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 147–164). New York: Macmillan.

Franklin C. A., & Garfield, J. (2006). The GAISE project: Developing statistics education guidelines for grades pre-K-12 and college courses. In G.F. Burrill and P.C. Elliott (eds) *Thinking and Reasoning with Data and Chance: 2006 NCTM Yearbook*, pp. 345–376. Reston, VA: National Council of Teachers of Mathematics.

Garfield, J. (2000). Assessment in statistics education: Issues and challenges. *Mathematical Thinking and Learning, 2*, 99-125.

Garfield, J. (2003). Assessing statistical reasoning. *Statistics Education Research Journal, 2*, 22-38. Retrieved April 25, 2009 at http://fehps.une.edu.au/serj.

Garfield, J. & Chance, B. (2000). Assessment in statistics education: Issues and challenges. *Mathematical Thinking and Learning, 2*, 99-125.

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*, 372-396.

Glaser, R., Lesgold, A., and Lajoie, S. (1987). Toward a cognitive theory for the measurement of achievement. In R. R. Ronning, J. Glover, J. C. Conoley, and J. C. Witt (Eds.), *The influence of cognitive psychology on testing and measurement* (pp. 41-85). Hillsdale, NJ: Erlbaum.

Huff, K., & Goodman, D. P. (2007). The demand for cognitive diagnostic assessment. In J. P. Leighton & M. J. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications* (pp. 19-60). Cambridge, UK: Cambridge University Press.

Jolliffe, F. (1991) Assessment of the understanding of statistical concepts, in D. Verejones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics.* Otago: Otago University Press.

Jones, G.A., Langrall, C.W., Mooney, E.S. & Thornton, C.A. (2004). Models of development in statistical reasoning.In *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking*, Eds. D. Ben-Zvi and J. Garfield, pp. 97–117. Dordrecht: Kluwer Academic.

Jones, G. A., Thornton, C. A., Langrall, C.W., Mooney, E. S., Perry, B., & Putt, I. A. (2000). A framework for characterizing children's statistical thinking. *Mathematical Thinking and Learning, 2*, 269–307.

Lalonde, R. N., & Gardner, R. C. (1993) Statistics as a second language? A model for predicting performance in psychology students. *Canadian Journal of Behavioral Science*, 25, pp. 108–125.

Leighton, J. P. (2004). Avoiding misconceptions, misuse, and missed opportunities: The collection of verbal reports in educational achievement testing. *Educational Measurement: Issues and Practice, 23*, 6-15.

Leighton, J. P., & Gierl, M. J. (2007a). Why cognitive diagnostic assessment? In J. P. Leighton & M. J. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications* (pp. 3-18). Cambridge, UK: Cambridge University Press.

Leighton, J. P., & Gierl, M. J. (2007b). Defining and evaluating models of cognition used in educational measurement to make inferences about examinees' thinking processes. *Educational Measurement: Issues and Practice, 26*, 3-16.

Leighton, J. P., & Sternberg, R. J. (2003). Reasoning and problem solving. In A. F. Healy & R. W. Proctor (Vol. Eds.), *Experimental psychology* (pp. 623-648). New York: Wiley.

Mislevy, R. J. (2006). Cognitive psychology and educational assessment. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 257-305). Washington, DC: American Council on Education.

Mooney, E.S. (2002). A framework for characterizing middle school students' statistical thinking. *Mathematical Thinking and Learning, 4*(1), 23–63.

Mooney, E.S., Langrall, C.W., Hofbauer, P.S., & Johnson, Y.A. (2001). Refining A Framework On Middle School Students' Statistical Thinking. In Speiser, R; Maher, CA & Walter, CN (Eds) *Proceedings Of The Twenty-Third Annual Meeting Of The North American Chapter Of The International Group For The Psychology Of Mathematics Education.* Vol 1: 437-447.

Onwuegbuzie, A. J. & Seaman, M. (1995). The effect of time and anxiety on statistics achievement, *Journal of Experimental Psychology*, 63, pp. 115–124.

Resnick, L. (1983). Toward a cognitive theory of instruction. In S. G. Paris, G. M. Olson, & H. W. Stevenson (Eds.), *Learning and motivation in the classroom* (pp. 5–38). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Sternberg, R. J, & Pretz, J. E. (Eds.). (2005). *Cognition and intelligence: Identifying the mechanisms of the mind.* New York: Cambridge University Press.

Zeidner, M. (1991) Statistics and mathematics anxiety in social science students—some interesting parallels, *British Journal of Educational Psychology*, 61, pp. 319–328.