



What Does Research Suggest About the Teaching and Learning of Introductory Statistics at the College Level? A Review of the Literature

Andrew Zieffler, Joan Garfield, Shirley Alt, Danielle Dupuis, Kristine Holleque, and Beng Chang
University of Minnesota

Journal of Statistics Education Volume 16, Number 2 (2008),
www.amstat.org/publications/jse/v16n2/zieffler.html

Copyright © 2008 by Andrew Zieffler, Joan Garfield, Shirley Alt, Danielle Dupuis, Kristine Holleque, and Beng Chang all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the authors and advance notification of the editor.

Key Words: Statistics Education Research; Teaching and learning; College students.

Abstract

Since the first studies on the teaching and learning of statistics appeared in the research literature, the scholarship in this area has grown dramatically. Given the diversity of disciplines, methodology, and orientation of the studies that may be classified as "statistics education research," summarizing and critiquing this body of work for teachers of statistics is a challenging and important endeavor. In this paper, a representative subset of studies related to the teaching and learning of statistics in introductory, non-calculus based college courses is reviewed. As a result of this review, and in an effort to improve the teaching and learning of statistics at the introductory college level, some guidelines to help advance future research in statistics education are offered.

1. Introduction

Since the first studies on the teaching and learning of statistics at the college level appeared in the research literature, this area of inquiry has grown dramatically (e.g., [Becker, 1996](#), [Garfield and Ben-Zvi, 2007](#)). Because of the diversity in backgrounds of researchers (e.g., psychology, educational psychology, statistics, statistics education, and mathematics education), scholarship related to the teaching and learning of statistics has appeared in journals from many different disciplines (e.g., *Journal for Research in Mathematics Education* [JRME], *Journal of Statistics Education* [JSE], *Statistics Education Research Journal* [SERJ], *Psychological Bulletin*, *Teaching of Psychology*). In addition, brief summaries of research related to teaching and learning statistics have been published in the journals *Teaching Statistics* and *Mathematics Teacher* and in conference proceedings from international conferences such as the International Conference on Teaching statistics (ICOTS) and education sessions at the International

Statistical Institute (ISI) (see <http://course1.winona.edu/cblumberg/books.htm>).

The more recently established journals in statistics education (JSE, SERJ) want to attract researchers from various disciplines whose work relates to teaching and learning statistics. However, in reviewing articles by researchers in disciplines such as mathematics education and psychology, it becomes apparent that many authors cite references published in their own disciplines, seemingly unaware of applicable studies in other disciplines or in the new statistics education journals. It seems statistics education is a new and emerging discipline. Although it is gaining some attention in the statistics world, often it is not recognized as a unique area of inquiry (Garfield, and Ben-Zvi, in press). For example, there is a lack of graduate programs and courses to prepare researchers in statistics education. Therefore research on the teaching and learning of statistics remains disconnected, fragmented, and difficult to access.

As an interdisciplinary field of inquiry, statistics education research has not relied on any one tradition of empirical research methodology. Instead, there have been a variety of research questions, methodologies, and outcome variables studied. Participants in the studies have ranged from young children to high school and college students to research professionals. Therefore, the focus of this paper is to review a subset of the research: a sample of studies related to the teaching and learning of statistics in introductory college courses. These studies typically take place in the non-calculus based statistics courses in the United States taught by mathematicians, statisticians, and psychologists.

The papers reviewed in this article were read, discussed and critiqued as part of a graduate research seminar held at the University of Minnesota in spring 2006. Five graduate students and one instructor wrote this paper together with the intent of providing a critical summary and analysis of a representative sample of research followed by suggestions for future research. The intent was not to read and review all published research on this topic, but to give a sense of the types of studies that have been done and what they have to offer.

This article begins with a general examination of the field of statistics education research, and then groups the studies reviewed into the following four categories: 1) Studies that identify misconceptions, faulty intuitions, and errors in reasoning; 2) studies that focus on assessing cognitive outcomes (what students learn in a statistics class); 3) studies that assess and study non-cognitive outcomes (e.g., attitudes); and 4) studies focused on the teaching of statistics at the college level. Each section concludes with a look at the limitations and practical implications of the studies for college teachers of statistics. Again, rather than provide a comprehensive set of suggested actions and activities, a description of some possible implications is offered.

2. What is Statistics Education Research?

Raudenbush (2005) has argued that improving instruction should be the key goal in any educational research. Therefore, the authors view the goal of statistics education research as the improvement of teaching statistics, leading to improved student learning. The Research Advisory Board of the *Consortium for the Advancement of Undergraduate Statistics Education* (CAUSE) describes statistics education research as being designed so the results will have direct implications for instruction, suggesting research studies in this area should specifically address classroom implications as well as the generation of new research questions (<http://www.causeweb.org/research/>). This definition serves as an ideal or model. However, in reviewing studies for this paper, many do not fit this ideal and indeed, many of the researchers would not define their work as statistics education research. We have taken a broad and pragmatic view of statistics education research for this paper that includes research studies that inform our understanding of the learning or teaching of statistics, believing that ultimately, this knowledge may lead to improvements in educating students in this area.

In reviewing publications that fit the category of statistics education research, it is apparent that there are different theoretical and research backgrounds for the studies, different research methods as well as different foci for the studies. Summarizing and critiquing this entire body of work for teachers of statistics is a challenging and important endeavor but is beyond the scope of this paper. Rather than provide an exhaustive summary, the authors have selected a few representative studies from a few broad areas that emerged from a review of the literature. The studies included in this paper are evaluated through the lens of how well their results inform the field of statistics education and meet the goals of statistics education research. As a result of this review, the authors, in an effort to improve the teaching and learning of statistics at the college level, offer some guidelines to help advance future research in statistics education.

3. Studies on the Nature and Cause of Faulty Statistical Reasoning

Studies primarily conducted by psychologists have revealed difficulties students and adults encounter when reasoning about chance events. The focus of these studies tends to be on probability, randomness, and their connection to uncertain outcomes. These studies are grouped into two categories: studies that primarily examined whether people exhibited faulty reasoning, and studies that attempted to understand why these types of errors prevail.

3.1 Studies that Explore the Nature of Faulty Reasoning

Beginning in the 1970's, several psychologists became interested in the faulty reasoning exhibited by adults in reasoning about problems involving uncertainty. The research produced by these researchers has helped to identify many misconceptions and errors in reasoning that are widespread and therefore relevant for teachers of statistics (e.g., [Garfield and Ahlgren, 1988](#); [Shaughnessy, 1992](#); [Garfield, 1995](#)).

The studies on misconceptions, faulty intuitions, and errors in reasoning about concepts in statistics have been greatly influenced by the pioneering work of [Kahneman, Slovic, and Tversky \(1982\)](#). They were among the first to discover a number of systematic and persistent errors people make when attempting to make decisions involving chance and uncertainty. These psychologists and their colleagues described and labeled several errors in reasoning, including:

- *Representativeness Heuristic*: when people estimate the likelihood of a sample based on how closely it resembles the population. Use of this heuristic also leads people to judge small samples to be as likely as large ones to represent the same population.
- *Base-rate fallacy*: when people ignore the relative sizes of population subgroups when judging the likelihood of contingent events involving the subgroups, especially when empirical statistics about the probability are available (called the "base rate").
- *Conjunction fallacy*: when people judge the conjunction of two correlated events to be more likely than either of the events themselves.

This research distinguished between *errors of application* (an individual knows a rule and fails to apply it) and *errors of comprehension* (an individual does not recognize the validity of the rule they violated) for the purposes of diagnosing the nature of errors people make. Based on students' justifications to specific problems, Kahneman, Slovic and Tversky (1982) proposed that some errors and biases in judgment under uncertainty call for a dual analysis that both explains the choice of a particular erroneous response in terms of heuristics and also explains why the correct rule has not been learned.

[Gigerenzer \(1996\)](#) criticized Kahneman, Slovic, and Tversky's findings stating that some of the cognitive processes were overlooked and misclassified. He reframed Kahneman, Slovic, and Tversky's research

questions in terms of frequencies and found that college students no longer made the same mistakes when reasoning about uncertain events. Other researchers, however, have replicated some of Kahneman, Slovic, and Tversky's findings (e.g., [Fong, Krantz, and Nisbett, 1986](#); [Nisbett, Fong, Lehman, and Cheng, 1987](#)). These researchers also conducted studies using university students and found that training them in formal aspects of the rules (e.g., teaching the rule by means of examples) improved both the frequency and quality of their statistical reasoning.

Psychology research in other areas of statistical reasoning has also produced inconsistent findings. For example, [Sedlmeier and Gigerenzer \(1997\)](#) examined the seemingly inconsistent research results on whether people take sample size into account in their reasoning about distributions. They proposed that one reason for the inconsistencies in the findings from this body of evidence was that the studies were using two kinds of tasks: one for which peoples' intuitions about large sample size are often correct (frequency distributions - where a large sample leads to a better representation of a population) and one for which this intuition may be misleading (sampling distributions - when a larger sample size leads to less variability between samples). Drawing from the results from several studies, they tentatively suggest that the intuitive reliance on the law of large numbers leads participants to construct frequency distributions even if they were asked to construct sampling distributions.

Another explanation for the divergent findings on people's probabilistic reasoning may involve the inferences based on specific words (e.g., how a person infers the meaning of terms such as "probable" from the context of the problem in question, or from the broader context of human communication) people make when solving statistics problems. For instance, [Gigerenzer \(1996\)](#) claims people are drawing on a variety of these semantic inferences to make a judgment, and that the meaning of "probable" is not reducible to a single rule. This explanation has also been supported by the study of university students conducted by [Sides, Osherson, Bonini, and Viale, \(2002\)](#).

3.2 Studies that Explore Why Students have Faulty Reasoning

Researchers have also studied how and why college students appear to have faulty reasoning and poor understanding of statistics. [Pollatsek, Lima, and Well \(1981\)](#) examined college students' understanding of the mean. Student interviews suggested that statistics students have only a functional (i.e., computational) concept of the mean. [Konold, Pollatsek, Well, Lohmeier, and Lipson \(1993\)](#) also studied students' faulty reasoning by examining college students' understanding of independence. Their interviews of college students showed that these students were using the outcome approach ([Konold, 1989](#)) – basing their prediction on data that are deterministically or causally linked to the event of interest. Metz (1998) took this exploration of uncertainty and corresponding forms of probability a step further and explored the complex issue of developmental versus non-developmental difficulties in students' understanding. Her research suggested that enduring challenges in the interpretation of random phenomena (e.g., failure to recognize whether uncertainty enters the situation) were manifested in kindergarten and continued through to adulthood.

3.3 What these Studies have to Offer Teachers of Statistics

The studies on faulty reasoning provide evidence that there are many errors and uses of faulty reasoning which can be quite difficult to correct in spite of good instructional methods and activities. These errors tend to be domain specific and dependent on the students' own personal and contextual interpretations. Statistics instructors who hope to improve students' statistical reasoning are advised to be aware of some possible types of errors and misconceptions. This may be done by using informal or formal methods of assessment to help them differentiate between students' correct and faulty reasoning in the classroom. Errors may then be addressed by different methods (e.g., particular activities using a simulation tool or web applet) to help build correct reasoning. For example, students could be asked to run a simulation of

coin tossing and their attention could be directed to runs of heads and tails to address the Gambler's Fallacy, an application of the Representativeness Heuristic that leads them to believe that after a long run of heads (when repeatedly tossing a fair coin) a tail is most likely on the next toss.

3.4 Questions and Considerations Raised by the Studies in this Area

While the set of studies concerning misconceptions and faulty reasoning created a new and prolific area of research, there have also been critiques of these studies, which led to additional studies exploring why these misconceptions and faulty reasoning exist. For example, some have questioned whether there really are prevalent misconceptions or whether people reason differently when provided with different contexts. Another issue is that many of these studies were conducted using paper and pencil, multiple-choice instruments and that reasoning was not probed using more intensive methods such as clinical interviews. Studies in this area that involved a training component and produced positive effects on overcoming misconceptions may be limited in their ability to generalize to educational settings (e.g., a statistics course) due to the artificial nature of the training in a lab setting. These studies may also be limited by their use of assessment to determine the effects of training, such as using a few items that correspond to the exact types of problems used in the study.

4. Studies on Students' Statistical Reasoning

While early studies focused primarily on the prevalence of faulty reasoning and misconceptions, more recent studies have examined the existence and development of reasoning about important statistical ideas. Both formal and informal methods of assessment have been used by statistics education researchers to study college students' statistical reasoning. Studies that focus on assessing students' statistical reasoning have typically revealed that even after formal instruction, many people remain unable to reason correctly about important statistical ideas and concepts (e.g., [Konold, Pollatsek, Well, Lohmeier, and Lipson, 1993](#); [Schau and Mattern, 1997](#); [Hirsch and O'Donnell, 2001](#); [Garfield, 2003](#)). Statistics education researchers have used a variety of methods to study college students' statistical reasoning.

4.1 Quantitative Approaches to Assessing Statistical Reasoning

One instrument that has been used in several studies of college students' statistical reasoning is the Statistical Reasoning Assessment (SRA; [Garfield, 1998a](#)). The SRA contains 20 multiple-choice items designed to assess students' ability to reason with statistical information (e.g., correctly interpreting probability, understanding independence and sampling variability, distinguishing between correlation and causation). Responses to each item include a statement of reasoning, explaining the students' rationale for their particular choice. The items are coded according to both the types of errors they assess as well as correct types of reasoning. The SRA has been used in different contexts and reasonable test-retest reliability and content validity have been established ([Garfield, 1998b, 2003](#); [Liu, 1998](#); [Garfield and Chance, 2000](#)). The results from different administrations show surprisingly similar (and poor) results, despite country or type of course in which the SRA was given. In a recent study, [Tempelaar, Gijsselaers, and Schim van der Loeff \(2006\)](#) examined the puzzle of low or zero correlations between aggregated reasoning skills and course performance that was found in studies using the SRA and other measures of statistical reasoning. Their study suggested that the puzzle may be at least partly understood in terms of differences in the effort students invest in studying: Students with strong effort-based learning approaches (e.g., many homework assignments, etc.) tended to have lower correct reasoning scores, and higher misconception scores, than students with different learning approaches.

[Hirsch and O'Donnell \(2001\)](#) created a 16-item measure that utilized both multiple-choice and open-ended items. Results of their analysis revealed that even after formal instruction in probability, a

substantial proportion of students continued to hold misconceptions, especially concepts surrounding representativeness. [Konold et al. \(1993\)](#) assessed students' consistency in reasoning using forced choice items that are included in the SRA. The results of their analysis reveal that students do not reason consistently across these items. The authors suggest that a possible explanation for this inconsistency is that students switch between two heuristics when reasoning about probability: the outcome approach and the representative heuristic. Konold's earlier work ([1989](#)) defined an outcome approach as one where the goal in questions of uncertainty is to predict the outcome of an individual trial in the form of yes-no decisions on whether an outcome will occur on a particular trial.

A more recently developed instrument is the Comprehensive Assessment of Outcomes in a First Statistics course (CAOS), developed as part of the ARTIST project (Assessment Resource Tools for Improving Statistical Thinking (<https://app.gen.umn.edu/artist/index.html>)). CAOS is a forty-item test that was designed to evaluate student attainment of desired outcomes in an introductory statistics course. The 40 multiple-choice items focus on the big ideas and "the types of reasoning, thinking and literacy skills deemed important for all students across first courses in statistics" ([Garfield, delMas, and Chance](#)). The CAOS test has gone through an extensive development, revisions, and validation process including class testing and reliability analyses (see [delMas, Garfield, Ooms and Chance, in press](#)).

4.2 Qualitative Approaches to Assessing Statistical Reasoning

In addition to the studies that developed and used written tests to assess students' statistical reasoning there have been several other types of studies that used more informal methods of assessment. These have included student interviews or the examination of students' responses to open-ended assessment items.

In a qualitative study, [Groth and Bergner \(2005\)](#) examined the use of metaphors to assess students' understanding of statistical ideas. In a pre-service methods course for teachers, they asked their students to write down their own metaphor for the concept of a sample and to discuss the strengths and limitations of the metaphor. Building on the framework developed by [Watson and Moritz \(2000\)](#), [Groth and Bergner](#) categorized the students' responses using the Structure of Observed Learning Outcomes model (SOLO; [Biggs and Collis, 1982](#)). The SOLO taxonomy provides a hierarchy of five different levels of subject matter understanding for evaluating student work. [Groth and Bergner](#) found that despite successful completion of an introductory statistics course, many of the students' conceptions of sample seemed limited to "a group of objects or a collection of numerical data (p. 38)".

In a set of studies conducted across four different universities, [Mathews and Clark \(2003\)](#) and [Clark, Mathews, Kraut, and Wimbish \(1997\)](#) used clinical interviews to assess student knowledge of the mean, standard deviation and the Central Limit Theorem. Student responses were classified using APOS (Action, Process, Object, Schema) theory ([Dubinsky and McDonald, 2001](#)). Most students were found to have relatively unsophisticated understandings of the concepts of mean and standard deviation and only fragmentary recall, at best, of the Central Limit Theorem. These studies suggest that even high performing students may not be able to reason about even basic statistical concepts, such as the mean and standard deviation, much less more abstract concepts, such as the Central Limit Theorem.

4.3 What these Studies have to Offer Teachers of Statistics

The assessment instruments and methods described in this section have provided additional data that demonstrate students' difficulties reasoning about probability and statistics. The assessment data reveal that even after formal instruction, many students do not reason consistently or accurately. Furthermore, the studies suggest an inconsistency between what students are able to demonstrate on homework, quizzes and exams and their ability to reason about statistics. This may be due to the fact that much of statistical reasoning involves rather complex and at times abstract concepts, and more traditional methods

of classroom assessment do not evaluate these higher-level cognitive outcomes. Difficult and complex concepts may need to be revisited multiple times throughout a course and developed carefully from informal, intuitive foundations to the more formal and abstract form of these concepts (see [Garfield and Ben-Zvi](#), in press, for examples of how to do this). A related concern is that teachers of statistics take care not to overestimate their students' reasoning skills, even if their students have performed well in an introductory course. It may also be worthwhile for teachers to carefully review and reflect on the types of tasks used in assessment of students' reasoning. For instance, having students explain their reasoning as they work in pairs or small groups on a classroom activity is one way to informally assess students' statistical reasoning. The types of tasks used by [delMas and Liu \(2005\)](#) about what makes the standard deviation larger or smaller can be used to promote this type of discussion.

4.4 Questions and Considerations Raised by the Studies in this Area

There are only a few quantitative instruments that are available to use in research studies on students' statistical reasoning. Looking at these instruments together reveal that a majority of the items focus on probability and fewer items assess understanding of statistical concepts. Given that there is more research (and a history of research) on probabilistic reasoning, this may not be surprising. However, more validated multiple-choice items are needed to assess reasoning about statistical concepts such as distribution, variability, and center. Another issue is that statistical reasoning may not be a single attribute or trait that can be measured in one dimension. Inconsistencies have been noted in how students reason across items and factor analyses (e.g., of the SRA) have not found a single factor or even a few meaningful sub-factors. It may be more useful to have several different assessments, each evaluating reasoning about a particular statistical concept (such as sampling). The use of qualitative methods should be further explored, and may offer ways to develop quantitative instruments based on questions used in the qualitative studies.

5. Studies of Non-cognitive Outcomes and Factors that Affect the Learning of Statistics at the College Level

Students' affect and attitudes are two of the "non-cognitive" aspects of statistics instruction that have been studied by statistics education researchers. These include beliefs and feelings about statistics and its utility; the student's perceived ability to successfully complete statistical tasks or solve problems; and feelings about the level of difficulty of statistics ([Gal and Ginsburg, 1994](#)). As more attention has been paid to college students' difficulty in learning statistics, several researchers have explored how these non-cognitive factors may be related to students' understanding and reasoning about statistical information.

Several of these studies are described in this section. These studies have often examined the role these non-cognitive outcomes, such as affect, play in helping college students succeed in an introductory statistics course. For example, in a recent study, [Budé, Van De Wiel, Imbos, Candel, Broers and Berger \(2007\)](#) developed and used a questionnaire to examine motivational constructs and their effect on students' academic achievement within a statistics course. They identified dimensions of causal attributions, outcome expectancy, affect, and study behavior, with respect to statistics. Their results of a path analysis found a relationship between negative attitudes toward statistics and poor study habits, which led to poor scores on achievement measures. Before these studies are reviewed, a sample of the instruments used to assess these non-cognitive factors is described.

5.1 Instruments to Assess Non-cognitive Factors Related to College Students Learning Statistics

Many different instruments have been used to study the affect and attitudes of introductory statistics

students. Several of these instruments are described in this section. Two instruments designed to examine self-efficacy of statistics students are the *Current Statistics Self-Efficacy* (CSE; <http://www.stat.auckland.ac.nz/~iase/cblumberg/finney2.doc>) and the *Self-Efficacy to Learn Statistics* (SELS; <http://www.stat.auckland.ac.nz/~iase/cblumberg/finney1.doc>). These two self-efficacy scales ask the examinee to rate their confidence about a concept-based task. Finney and Schraw (2003) state the difference between the two scales lies in their instructions. Respondents on the CSE are asked to rate their "current ability to successfully complete the following tasks," while the SELS asks examinees to rate their "confidence in learning the skills necessary." Both the CSE and SELS have been validated and their scores have been supported by rigorous reliability analyses.

The *Survey of Attitudes Toward Statistics* (SATS; <http://www.unm.edu/~cschau/downloadsats.pdf>), developed by Schau ([Schau, Stevens, Dauphinee, and Del Vecchio, 1995](#)), the *Attitude Toward Statistics Scale* (ATS; <http://www.stat.auckland.ac.nz/~iase/cblumberg/wise1.doc>), developed by [Wise \(1985\)](#), and the *Statistics Attitude Survey* (SAS) developed by [Roberts and Bilderback \(1980\)](#) are three assessments used to measure college students' attitudes about statistics. These attitude scales consist of statements about statistics that students' rate on a Likert-type scale. These instruments have demonstrated acceptable psychometric properties under the rationale of classical test theory (e.g., [Cashin and Elmore, 2005](#); [Schau, Stevens, Dauphinee, and Del Vecchio, 1995](#)).

5.2 Studies on the Role of Non-cognitive Factors in Learning Statistics

The research on the learning of statistics by college students has provided evidence supporting the affect of non-cognitive factors in learning statistics. According to [Finney and Schraw \(2003\)](#), self-efficacy concerning statistics plays an important role in not only students' attitudes about statistics, but also in influencing their performance in a statistics course. The potential role of statistics anxiety in particular has been well documented in the literature. [Onwuegbuzie and Wilson \(2003\)](#) define statistics anxiety as "anxiety which occurs when a student encounters statistics in any form and at any level (p. 196)" – in other words, anxiety and negative feelings elicited by statistics in general that may hinder performance on statistics-related tasks. They further assert that the research findings support statistics anxiety as one of the biggest predictors of achievement in research methodology and statistics courses. Gal and Ginsburg (1994) expand this idea by positing that statistics anxiety could engender ill-prepared students who may not master material because they feel hindered by nervousness.

5.3 Qualitative Studies on Non-Cognitive Factors

Several studies have focused on students' experience with statistics (e.g., [Reid and Petocz, 2002](#)) and on student reflections of their experiences and understandings of what it means to do statistics (e.g., [Gordon, 1995](#)). These studies often employ a qualitative methodology such as phenomenology - a philosophy-based approach used to understand a person's experience (see [Van Manen, 1990](#)) - that enables the researchers to provide rich descriptions and analyses of students' experiences, attitudes and reflections of their statistical endeavors.

[Gordon \(1995\)](#) examined student attitudes and expectations of statistics courses by analyzing student responses to interviews and surveys. She used Activity Theory – a philosophical framework in which human activities can be explained by the subjects' interaction with socially situated phenomena ([Leont'ev, 1981](#)) - as a lens to interpret her data, namely positing that students' perceptions, personal and societal goals, as well as their views on the relevance of statistics, influence their learning of the subject matter. More specifically, a student who sees statistics as a barrier to a chosen vocational path might not value being asked to strive for depth in understanding statistical notions.

[Reid and Petocz \(2002\)](#) asked their student participants to reflect on a shared experience of taking

undergraduate statistics courses. The students' responses were found to follow a hierarchy of views, from a conception of statistics as comprising unconnected numerical tools and procedures to views of statistics as modeling, as well as explaining data and worldly phenomena. The authors provide evidence that first year and third year undergraduates exhibited the whole range of views, suggesting that even students who take more statistics courses may not necessarily understand the purposes and work of statistics at a higher level.

5.4 What these Studies have to Offer Teachers of Statistics

The studies of non-cognitive factors influencing college students' learning of statistics have important implications for teachers of statistics. For example, careful use of instruments such as the SATS, ATS, SAS, CSE and SELS may allow teachers to learn about students attitudes and anxiety about learning statistics; what value judgments students place on the study of statistics; how difficult they perceive statistics to be; and how useful they perceive statistics to be. However, if a teacher is interested in using these instruments to measure pre-post course change, it is important to note that affective components measured using many of these instruments may be quite stable from the beginning to the end of an introductory course (e.g., [Gal, Ginsburg, and Schau, 1997](#)) and as Gal and Ginsburg (1994) remind us, "students'...affective responses or self-confidence regarding statistics, may be more labile and likely to fluctuate depending on changing circumstances and classroom events. Thus, interpretation of score changes needs to take into account the expected stability over time of the constructs being measured."

Teachers need to recognize that it is important to try to make the study of statistics a positive one for students and to find ways to bring in examples of interesting research studies, data sets, and problems. In particular, it is important that teachers of statistics remain aware that students come to statistics courses with great variation in expectations and perceptions of what statistics is about. It is recommended that statistics teachers help students experience the practice of statistics, which in turn helps them to understand its power and utility. This could involve having students work with data – from collection to analysis – to answer genuine and interesting statistical questions. [Reid and Petocz \(2002\)](#) make the argument that an awareness of the professional statistician's point of view is important for helping students to appreciate why certain concepts are worthy of learning. Additionally, by trying to improve students' attitudes toward statistics, students may be more motivated to attend class and engage in activities and challenging tasks, which could ultimately improve their learning of statistics.

5.5 Questions and Considerations Raised by the Studies in this Area

While there is a fair amount of research on the non-cognitive factors that are believed to influence college students' learning of statistics, there is a lack of empirical evidence to support that these factors are actually related to students' learning. Moreover, since many of these studies only attempt to characterize students' perspectives on statistics (e.g., [Gordon, 1995](#); [Reid and Petocz, 2002](#)), the recommendations for classroom teachers must remain vague generalities (e.g., communicating to students the importance of statistics to their lives). These types of interventions appear somewhat limited in light of the research that suggests students' attitudes do not appear to change from the beginning to the end of a course in introductory statistics, despite interventions focused on changing them (e.g., [Gal, Ginsburg, and Schau, 1997](#)). Lastly, [Hembree \(1988\)](#), in a meta-analysis of 562 educational studies, suggested that student attitudes often correlate significantly and negatively with general test anxiety. This finding could lead to the potential conclusion that anxiety toward statistics may stem from a general anxiety toward performance evaluation.

6. Research Studies Focused on Teaching Statistics to College Students

Given the amount of research that has documented the many difficulties students incur when reasoning about statistical information, it is not surprising that many researchers have focused their effort on how to improve student learning through classroom interventions (e.g., better pedagogy). These studies have taken place across disciplines, focusing on a variety of different student populations taking introductory statistics courses, and have investigated many different research questions.

6.1. Studies focused on Student Learning

Several researchers have studied problems related to student learning of statistics in college courses. For example, [Derry, Levin, Osana, Jones and Peterson \(2000\)](#) explored the use of small-collaborative activities to stimulate complex, real-life problem solving. They evaluated and interviewed college students who were completing a statistical reasoning course, and they observed evidence of pre- to post-course improvements in the quality of students' statistical reasoning. However, interviewing the students also revealed pervasive confusion in the students' thinking. One specific problem documented by these researchers was students' confusion between random sampling and random assignment, which adversely affected their ability to suggest ways to evaluate and improve flawed research designs.

It has been further suggested in the literature that student difficulty may stem from their inability to recognize structural similarities between statistics problems. A study by [Quilici and Mayer \(2002\)](#) tested the question of whether structural awareness could be taught. They conducted a within-subjects comparison of students before and after they took their first statistics course. They found that both schema instruction and examples significantly increased students' sensitivity to the structural properties of statistics word problems compared to a control group. Their results support a structural awareness theory in which students can be taught to form problem schemas by abstracting the underlying structural features of a problem statement and organizing them into a generalized problem model.

With the abundance of technology now available for use in teaching statistics, the question arises as to whether or not these tools can be used to improve students' statistical reasoning abilities. [Lane and Tang \(2000\)](#) compared the effectiveness of simulations for teaching statistical concepts to the effectiveness of a textbook. They found that students randomly assigned to the simulation training program performed better than those assigned to the traditional textbook approach and further, that formal statistical training transferred to problems of every-day reasoning. This finding is somewhat inconsistent with the findings of [Aberson, Berger, Healy, Kyle and Romero \(2000\)](#). They compared the performance of students who were randomly assigned to either a web-based interactive tutorial group or a lecture group and found that both groups improved from pretest to post-test with no statistically significant differences between the groups.

[Lovett \(2001\)](#) collected participants' talk-aloud protocols and synchronized them with computer traces of their interactions with the statistics package to find out what ideas and strategies they were using to solve data analysis problems. She assigned students to one of two conditions (end-only and immediate feedback) and then asked them to complete four experimental phases (pretest, instruction, problem solving and post-test) with one condition receiving more or less feedback as they worked through each problem. She found that students (with feedback) improved in their ability to select appropriate data displays, showing that a set of sub-skills involved in selecting appropriate analyses could be reasonably well learned in a short, focused lesson that forces students to practice making these selections on their own.

But is it possible to teach the "big picture" to students and improve their statistical reasoning skills? [Meyer and Lovett \(2002\)](#) believe so. They developed a computerized cognitive tutor (*SmartLab*) in which students solve data-analysis problems and receive individually tailored feedback. Study participants watched taped videos of lectures and worked on problems for the first four sessions, and then in the fifth

session worked on additional open-ended data-analysis problems on the computer while receiving minimal scaffolding. Participants' post-test (paper and pencil) scores increased significantly. These results suggest that when participants encountered new problems they were thinking about them in terms of the appropriate analysis rather than in terms of subject matter – they seemed to be getting the big picture. Researchers have found that use of specific technology tools can help illuminate difficulties in understanding particular concepts such as sampling distributions ([Chance, delMas and Garfield, 2004](#)) and standard deviation ([delMas and Liu, 2005](#)).

6.2 Studies focused on Pedagogical Methods

Some statisticians who teach statistics at the postsecondary level have also focused their scholarship on the teaching and learning of statistics. Some of these statistics educators have investigated factors that may affect student outcomes, such as a "new" or different instructional method. This is typically done in their own course or in multiple sections of a course at a single institution.

The most common factor studied is the use of a particular type of instruction. For example, [Keeler and Steinhorst \(1995\)](#), [Giraud \(1997\)](#) and [Magel \(1998\)](#) all investigated the use of different methods of cooperative learning in teaching statistics, and found positive results. [Keeler and Steinhorst](#) found that when students worked in pairs, their final grades were higher and retention of the course material was higher than in previous semesters. Giraud found that students in a class that used in-class cooperative groups to work on assignments had higher test grades than students in a lecture class. [Magel](#), who implemented cooperative groups in a large lecture class, also found improvement in students' test scores compared to previous semesters.

Some researchers have explored the use of different types of technology to help teach statistics. For example, [Alldredge and Brown \(2006\)](#) investigated the use of two different types of multimedia packages on course performance and found a complex relationship between gender, student beliefs, the type of instructional software, and achievement. However, a confounding variable in the study was the use of two different types of statistical software, each used with just one multimedia package. Another study explored the effects of using graphing calculators on students' performance on an exam about statistical inference ([Collins and Mittag, 2005](#)). The researchers found that although use of the calculator with inferential capabilities was associated with improved scores on the exam, the improvement was not significant once they adjusted for performance on previous tests. In fact, there were few differences in the difficulty students had with final exam items.

Another topic of interest to statistics educators has been the use of online instruction either in a web-based or hybrid course (that includes a combination of face-to-face and web-based instruction). For example, [Utts, Sommer, Acredolo, Maher, and Matthews \(2003\)](#) and [Ward \(2004\)](#) found no differences in course performance for students in a hybrid versus a traditional course. They concluded that hybrid courses were not resulting in decreased student performance, although Utts et al. noted lower evaluations by students in the hybrid courses. However, the fact that no significant differences were found does not imply that there were no real differences in student outcomes for the compared instructional methods. For example, both studies utilized instructor generated assessments as measures to compare course performance. And, while there were no differences found in student performance between the two types of courses *on these assessments*, there may be differences in important outcomes such as statistical reasoning that were not measured.

6.3. Studies Focused on Developing a Particular Concept or Type of Statistical Reasoning

Heavily influenced by studies in mathematics education, classroom based research on the teaching and

learning of statistics at the college level has been used to study the development of particular types of statistical reasoning using specific activities and tools. These studies have typically measured student learning during and after a specific set of activities or tools were introduced in the classroom.

[Meletiou and Lee \(2003\)](#) organized their curricula along a Project-Activities-Cooperative Learning-Exercises model emphasizing statistical thinking and reasoning as well as an orientation towards investigating conjectures and discovery of results using data. Utilizing a pre-post design, students were assessed on their understanding at both the beginning and end of the course. Increases in understanding were observed on tasks requiring statistical reasoning (e.g., deducing whether a set of data could have been drawn at random from a particular population).

Using a collaborative classroom research model in a multi-institution study, [delMas, Garfield, and Chance \(1999\)](#) examined the development of students' reasoning about sampling distributions using a simulation program and research-based activities. Using a post-test designed to assess students' reasoning about sampling distributions, they found that student performance improved as the students were asked to make and test conjectures about different empirical sampling distributions from various populations. [Lunsford, Rowell and Goodson-Espy \(2006\)](#) replicated this study in a different type of introductory undergraduate statistics course and found similar results.

Building on these collaborative classroom research methods, [Garfield, delMas and Chance \(2007\)](#) used Japanese Lesson Study (see [Fernandez, 2002](#)) to design, test, and revise a lesson to help students develop reasoning about variability. A group of teachers of introductory statistics courses designed a lesson based on research that they felt would help reveal and build on student's informal intuitions about variability. This lesson was subsequently taught, observed, analyzed, and revised. From the resulting lesson emerged a sequence of activities that might help students develop a deep understanding of the concept of variability. In a related study, delMas and Liu (2005) studied students' development of reasoning about standard deviation. By having students manipulate a software tool to create histograms with the highest or lowest possible standard deviation given a set of fixed bars, the authors identified some common ways in which students both understand and misunderstand the standard deviation.

In a recent study, [Zieffler \(2006\)](#) studied the development of students' reasoning about bivariate data throughout an introductory statistics course using linear mixed effects (LME) models. He found that most of the students' development of covariational reasoning occurs prior to the formal classroom instruction of a bivariate data unit. This perhaps resulted from students learning to reason about graphs and data early in the course. Zieffler recommended the further use of statistical models that are aligned with educational models of learning to study the development of different types of statistical learning outcomes over time. For example, one could use a LME model with a quadratic term to capture both the learning and forgetting of a concept over time.

6.4 What these Studies have to Offer Teachers of Statistics

While some of the recent studies focused on teaching statistics in college classrooms have tried to compare different instructional methods, often the results of these studies are limited to the particular courses involved in the study and cannot be generalized to other courses. Other classroom research studies, while not trying to compare teaching methods, have suggested some practical implications for teachers. For example, developing a deep understanding of statistics concepts is quite challenging and should not be underestimated. Research suggests that it takes time, a well thought out sequence of learning activities, appropriate tools, and discussion questions. Good reasoning about important statistical concepts can be developed very carefully using activities and tools given enough time and revisiting of these ideas.

6.5 Questions and Considerations Raised by the Studies in this Area

Many of these studies tended to examine broad research questions that seemed beyond the scope of a single course or classroom study. Regarding the posing of research questions, [Garfield \(2006\)](#) suggests that researchers "(k)eeep the focus narrow (p. 8)." For example, rather than asking "does technology improve student learning," instead ask how a particular technology helps students understand a particular statistical concept. Another consideration of this type of research is that studies typically use course specific student outcomes, such as final exam grades or course evaluations, as a dependent measure. Because of the singular dependence of these outcomes to a particular course, many of the research results from these studies lack any desirable external validity. Furthermore, it is often difficult to understand the learning outcomes of courses described in most of these research studies because there is an omission of the assessment items that were used by the researcher. So the reader does not know if the students were tested on computational and procedural skills, or on higher levels of thinking and reasoning? (Suggestions regarding some of these assessment concerns are further detailed in section 7.2.)

Another consideration in classroom research, especially when investigating particular teaching methods or technology tools is the difficulty in drawing valid conclusions about the impact of that method or tool. For example, cooperative learning was used quite differently in each of the three studies cited regarding cooperative learning. Thus, it may be hard to generalize about the value of this method beyond its particular use in those three specific classes being studied. For example, the types of technology used in the studies cited above were quite different. It is important to note that there are many different types of technology and depending on the way they are used, they may or may not benefit students' learning.

Yet another consideration for future empirical research studies in statistics education is a design that compares two courses or two sections of a course, where one is "traditional" and one implements a new intervention; for example, a study in which one section uses a graphing calculator and another section that uses no technology, or a study in which one section uses multimedia and one uses "traditional" lectures. With these types of designs there is no standard "traditional" course in terms of content or teaching method (although traditional courses are often described as being lecture based). Therefore, if the use of an intervention shows positive gains, all that can be gleaned is that the use of that intervention in that course resulted in higher performance, but those results may not necessarily generalize to other courses using this intervention when compared to other "traditional" courses. Alternatively, studies that examine the impact of particular activities or approaches on students' understanding of particular concepts or learning outcomes rather than achievement in general appear to be more informative (e.g., the impact of using simulation software on developing an understanding of sampling distributions). The authors do note, however, that once there are more empirical studies published on the use of particular pedagogical methods, a meta-analysis may be used to provide a better understanding of the effectiveness of a method on learning statistics. Meta-analyses of methods like cooperative learning in education, for example, have demonstrated the positive effects of this method in different educational settings and levels (e.g., [Roseth, Johnson, and Johnson, in press](#)).

7. Summary

This paper has examined the research on the teaching and learning of statistics at the college level. This research has been published in several journals that, not unlike the researchers themselves, come from varying academic backgrounds. Likewise, the types of research questions, interventions, and methodologies used in these studies were markedly different. In this section, the authors of this article look at the practical implications of all of this research for teachers of statistics at the college level. They also offer some general criticism of this body of research with an eye toward improving future research endeavors on the teaching and learning of statistics at the college level.

7.1 Practical Implications for Teachers

The research reviewed in this paper has several implications for statistics teachers at the college level. Many of the studies have shown that statistics remains a pervasively difficult subject for students. Several of the concepts that college students encounter in an introductory course (e.g., sampling distributions, power) have been extensively studied with disheartening results (e.g., [Chance, delMas, and Garfield, 2004](#); [Garfield and delMas, 1994](#)). Students have shown inconsistencies in their reasoning about the most "elementary" concepts of measures of central tendency and spread (e.g., [Chance, delMas, and Garfield, 2004](#); [delMas and Liu, 2005](#); [Groth and Bergner, 2006](#); [Noss, Pozzi, and Hoyles, 1999](#)). Research has also shown that many college students enter introductory statistics courses with a fair amount of trepidation and anxiety, and that these feelings tend to persist throughout the course (e.g., [Gal and Ginsburg, 1994](#); [Perney and Ravid, 1991](#)).

While these findings seem apocalyptic for students' opportunities to learn statistics, there is much that teachers can learn from these results. For starters, teachers should not overestimate the degree to which they believe students have learned or understand even the most basic concepts in an introductory statistics course. Teachers looking to find out what their students actually know can benefit from the use of "good" assessment items (see for example the Assessment Resource Tools for Improving Statistical Thinking; [Garfield, delMas, and Chance](#)) rather than using their own constructed items, which typically have less desirable psychometric properties (e.g., [Gullickson and Ellwein, 1985](#); [Weller, 2001](#)).

Another suggestion for teachers is to examine students' response patterns across assessment items to help identify inconsistencies in their statistical reasoning. Some studies on student reasoning (e.g., [Konold, 1995](#)) have found that patterns of responses can reveal surprises and inconsistencies in students' reasoning. For example, [delMas, Garfield, Ooms, and Chance \(in press\)](#) show how students can readily identify a correct definition of p-value, but many of these same students also identify an incorrect definition of p-value as correct. Looking at cross-tabulations of responses to items such as these can illuminate errors in student understanding and allow the instructor to address and confront some of these errors. While changing student affect seems to be a daunting task for any teacher of statistics, there is evidence to suggest that the anxiety students feel toward statistics may be related to mathematics anxiety (e.g., [Birenbaum and Eylath, 1994](#)) and that this may be reduced by de-emphasizing the mathematics that occurs in a more "traditional" course. Likewise, educational research shows that some test anxieties can be decreased by using more non-traditional types of assessments such as student projects or portfolios ([Lissitz and Schafer, 2002](#); [Reys, Lindquist, Lambdin, and Smith, 2006](#)).

Lastly, statistics teachers at the college level could benefit from reading the literature on the teaching and learning of statistics. Often, the research not only points out pitfalls for students in learning statistics, but also offers insight into possible remedies via technology or teaching points. The research literature can provide a blueprint for designing or improving curriculum. Research from several disciplines (e.g., learning theory, statistics education, mathematics education) can help in formulating both the scope and sequence of the introductory statistics curriculum, while guidelines for teaching statistics, such as those found in the Guidelines for the Assessment and Instruction in Statistics Education ([GAISE; American Statistical Association, 2005](#)), can help teachers implement many of the suggestions found in that research.

7.2 Three Specific Considerations for Researchers

After reviewing a selective sample of the literature available on the teaching and learning of statistics at the college level, three specific considerations to help improve future research conducted in this area are elaborated on: literature reviews, measurement issues, and external validity. This is done in an effort to help improve the quality of future research in this area. While these issues aren't the only considerations

to be made in a research study, they are a good place to start for future researchers.

Literature Reviews. The first suggestion relates to the review of the literature conducted and reported in these studies. The role of the literature review is to provide a critical review and analysis of the research relevant to the particular topic being studied. Literature reviews in the above reviewed studies primarily included research from the discipline in which the researcher is grounded. As an interdisciplinary area of research, statistics education researchers need to reflect that in their evaluation of the prior research. Examining the research from only one field of study (e.g., psychology) does not allow for a full picture for the researcher or the reader. This limited view of the prior research may make it difficult for readers (and even in some cases the researchers themselves) to see how the proposed study adds to the relevant research in the field.

Aside from thoroughly describing previous work done in the specific area of research, the literature review also needs to evaluate this work by knitting together the theories and results from the studies reviewed to describe the "big picture" of a field of research. Rather than offering only summaries of the studies included in the literature review, a synthesis of the prior research helps unify the research on a particular problem, which allows the literature review to operate as an organized and critical discussion, letting the reader see how the literature reviewed is relevant to the research question being examined. The synthesis of the available literature also helps to contextualize the research within the field by identifying where there are gaps in previous research that the study being proposed will help to fill.

Therefore, researchers are urged to thoroughly review the related literature and use this as a basis for framing their problem and reporting their results. While there are several different ways to search for articles (e.g., Educational Resources Information Center, <http://www.eric.ed.gov/>), one recommendation that is an excellent source for finding a vast array of articles across disciplines related to statistics education is the searchable literature index located on the CAUSE website (<http://www.causeweb.org/research/>). Using this index, one can search on statistical terms (e.g., correlation) as well as pedagogical methods (e.g., cooperative learning) to retrieve articles that directly relate to statistics teaching and learning on those topics.

Measurement Issues. Measurement refers to "the process of quantifying observations [or descriptions] about a quality or attribute of a thing or person (Thorndike and Hagen, 1986, p.5)." Thorndike and Hagen break this process into three steps: identifying and defining the quality or attribute that is to be measured; determining a set of operations by which the attribute may be made manifest and perceivable; and establishing a set of procedures or definitions for translating observations into quantitative statements of degree or amount (Thorndike and Hagen, 1986). For example, in many of the studies reviewed in this paper, students' statistical knowledge or reasoning was translated into a degree of quantification by the assignment of a test score to each student. These scores are then generally subjected to some kind of quantitative analysis.

Since the measurements used in research studies are essential to the findings that are produced, it is important that they are valid (measure the constructs they purport to measure) and reliable (have consistency, either over items in the test, or over time) (Pedhazur and Schmelkin, 1991). Descriptions of both the development of the measurements used in a study and evidence of their meaningfulness and appropriateness to the groups or participants being studied are essential elements in the reporting of research, especially in circumstances under which a knowledgeable scholar might reasonably have questions.

Many of the articles reviewed used final exams, course grades or percentages as a proxy for statistical knowledge or reasoning. However, there is evidence or argument that these outcome measures are not valid indicators of statistical knowledge or reasoning (e.g., Chance and Garfield, 2002; Konold, 1995).

Likewise, unless it is included with the article, the reader has no way of judging if a final exam primarily measures computational ability or conceptual understanding. Therefore, it is hard to interpret what scores on such a test are measuring. In addition, some items on a test may be poorly worded and may not be good measures of a student's understanding. Using final grades or percentages are also inadequate, because again, the reader does not know what is being assessed and valued as outcome measures. Sometimes authors include a copy of their exam or sample items for the readers to view, to give a sense of what has been measured. However, this judgment of whether the test appears to be a reasonable instrument to measure the stated outcomes (i.e., face validity) provides insufficient validity evidence ([Thorndike, 2004](#)).

A set of high quality instruments to assess important, agreed upon learning goals is needed for researchers to use in their studies. The ARTIST project has developed a few such instruments, such as the Comprehensive Assessment of Outcomes of a First Statistics Class (CAOS; <https://app.gen.umn.edu/artist/index.html>). Data are now available from a large number of examinees so that researchers have some normative data to use in comparing their results (see [delMas, Garfield, Ooms and Chance, in press](#)). The authors urge other researchers to consider the development and validation of assessment instruments as a research endeavor that can serve the statistics education community, whether or not it is part of a larger project.

External and Internal Validity. An experiment is said to have external validity if the results hold across different settings, procedures and participants, allowing results to generalize to the larger population. While a call for randomized studies in education might offer some evidence for internal validity – establishing a causal relationship through randomization of treatments (see [Mosteller and Boruch, 2002](#)) – there is still the problem of which population of students the findings may be generalized to. For example, several studies on comparing "traditional" courses to web courses have used randomization in one fashion or another. But the questions of generalizability still remain. For instance, what constitutes a "traditional" introductory statistics course? There is no operational definition of what a "traditional" course is. Likewise, are all web courses created equal (or for that matter "traditional courses")? These problems in external validity, even with true experiments, often don't provide the statistics education community with any real answers about "what works" at the end of the day. Another concern is that many times students and teachers taking part in research are volunteers, and this may have consequences for both the internal and external validity of research (e.g., [Gall, Borg, and Gall, 1996](#)).

Connecting these three considerations. It is important to have carefully read and reviewed the literature before planning a study; to select or design high quality measurement instrument;, and to make sure that the researchers have considered the extent of and threats to both external and internal validity, These principles are well described in a recent report developed and distributed by the American Statistical Association ([Scheaffer et al., 2007](#)). While the focus of this report is on reporting guidelines for educational research in mathematics, there are important implications for designing high quality studies in statistics education.

8. Suggestions and Recommendations for Future Research

Despite the critiques and limitations of educational research there is much to be optimistic about in the future for statistics education research. Research-based guidelines for teaching statistics (GAISE) have been endorsed by the American Statistical Association and there are plans to form collaborative groups focused on statistics education research through research clusters in the NSF-funded CAUSEmos Grant (CAUSE Making Outreach Sustainable for statistics educators) as well as a new multidisciplinary project at The Ohio State University: Institute for Quantitative Education Research Infrastructure (INQUERI). This type of national consortium could help propel the discipline forward by establishing a culture of

collaborative researchers that can access student- and teacher-level data from multiple institutions and courses. Both of these efforts will rely on a new set of guidelines for using statistics in mathematics education research that are equally applicable to statistics education research (see <http://www.causeweb.org/research/guidelines/>).

With the number of college students taking statistics classes continuing to increase ([Lutzer, Maxwell, and Rodi, 2000](#); [Cobb, 2005](#)), good research on the teaching and learning of statistics at the college level is now more pertinent than ever before. With that in mind, some suggestions and recommendations for future research in this area are offered.

While the studies on peoples' difficulties with probability and errors in heuristics stimulated much of the current research on the teaching and learning of statistics at the college level, these areas now seem saturated. With current recommendations and guidelines calling for an emphasis on statistical reasoning and thinking as student outcomes in statistics classes, research needs to shift to help explain these outcomes. For instance, despite many papers on the importance of statistical thinking as a learning goal (e.g. [Moore, 1998](#); [Snee, 1990](#); [Snee 1999](#)), [Wild and Pfannkuch \(1999\)](#) were among the first researchers to conduct empirical studies and develop a theory and model of statistical thinking. Researchers are encouraged to explore and develop models and theories to help explain not only the nature of these important learning outcomes, but also student development within them.

Another area of research is in the development of student assessments that provide valid and reliable measurements on important student outcomes such as statistical reasoning or thinking. In developing these instruments, it is important that they demonstrate the capacity for consistently measuring student reasoning (reliability) or thinking on specific statistical idea (validity). Development and use of these instruments would not only allow the research and teaching community to aggregate and compare information collected from different studies, but could also be used in singular studies by providing the statistics education community with an acceptable yardstick for measuring *real* student achievement after a particular teaching or curriculum intervention (see [National Research Council, 2001](#)).

The authors believe it would be helpful for the statistics education community to create and make available some guidelines or standards to further improve the research being conducted and published on the teaching and learning of statistics at *all* levels. The American Educational Research Association (AERA) has recently drafted a set of standards to aid researchers in reporting their research methods ([American Educational Research Association, 2006](#)). While these guidelines are aimed at helping researchers improve what they write about their research methods, their ultimate goal is to set a standard for the quality of research that is published and disseminated in AERA journals. Since there seems to be a lot of variation in the types and quality of studies and research in statistics education, a set of guidelines would be a major asset to the field. These guidelines would go a long way toward improving the overall quality of research and publication of research in the field, and thereby improve both the teaching and learning of statistics.

Finally, the authors encourage the forming of collaborative research groups, involving teachers of statistics as well as faculty from other disciplines such as psychology or education. See [Garfield and Ben-Zvi \(in press\)](#) for more arguments and suggestions for this type of research. [Garfield \(2006\)](#) offers additional reasons for collaboration: "Look for collaborators who share your research interests but who may bring different background (even disciplines) and strengths to a new collaboration (p. 8)." This removes the pressure of having to be an expert in everything.

While by no means offering a comprehensive review or critique of the research literature on teaching and learning introductory statistics at the college level, the authors have attempted to examine a representative sample of the literature available in this area. They have also provided suggestions, advice and

recommendations for researchers working in this area or who would like begin a study related to teaching and learning statistics. There is much about which to be optimistic regarding the future of statistics education research and the practical and insightful implications these studies may provide on the teaching and learning of statistics.

References

- American Educational Research Association. (2006), *AERA Draft Standards for Reporting on Research Methods*,
http://www.aera.net/uploadedFiles/Publications/AERA_Standards/AERASStandardsforReview.pdf
- American Statistical Association. (2005), *GAISE College Report*,
<http://www.amstat.org/education/gaise/GAISECollege.htm>
- Aberson, C. L., Berger, D. E., Healy, M. R., Kyle, D. J., and Romero, V. L. (2000), "Evaluation of an Interactive Tutorial for Teaching the Central Limit Theorem," *Teaching in Psychology*, 27(4), 289-291.
- Allredge, J. R., and Brown, G. R. (2006), "Association of Course Performance with Student Beliefs: An Analysis by Gender and Instructional Software Environment," *Statistics Education Research Journal* [Online], 5(1), 64-77. [http://www.stat.auckland.ac.nz/~iase/serj/SERJ_5\(1\)_Allredge_Brown.pdf](http://www.stat.auckland.ac.nz/~iase/serj/SERJ_5(1)_Allredge_Brown.pdf)
- Becker, B. J. (1996). "A Look at the Literature (and Other Resources) on Teaching Statistics," *Journal of Educational and Behavioral Statistics*, 21(1), 71-90.
- Biggs, J. B., and Collis, K. F. (1982), *Evaluating the Quality of Learning: The SOLO Taxonomy (Structure of Observed Learning Outcomes)*, New York: Academic Press.
- Birenbaum, M., and Eylath, S. (1994), "Who is Afraid of Statistics? Correlates of Statistics Anxiety Among Students of Educational Sciences," *Educational Research*, 36(1), 93-98.
- Budé, L., Van De Wiel, M. W. J., Imbos, T., Candel, M. J. J. M., Broers, N. J., and Berger, M. P. F. (2007), "Students' Achievements in a Statistics Course in Relation to Motivational Aspects and Study Behaviour," *Statistics Education Research Journal* [Online], 6(1), 5-21.
[http://www.stat.auckland.ac.nz/~iase/serj/SERJ6\(1\)_Bude.pdf](http://www.stat.auckland.ac.nz/~iase/serj/SERJ6(1)_Bude.pdf)
- Cashin, S.E., and Elmore, P.B. (2005), "The Survey of Attitudes Toward Statistics Scale: A Construct Validity Study," *Educational and Psychological Measurement*, 65(3), 509-524.
- Chance, B., dleMas, R., and Garfield, J. (2004), "Reasoning About Sampling Distributions," in *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*, eds. D. Ben-Zvi and J. Garfield, Dordrecht, The Netherlands: Kluwer Academic Publishers, pp. 295-323.
- Chance, B. L., and Garfield, J B. (2002), "New Approaches to Gathering Data on Student Learning for Research in Statistics Education," *Statistics Education Research Journal* [Online], 1(2), 38-41.
<http://www.stat.auckland.ac.nz/~iase/publications/4/751.pdf>
- Clark, J., Mathews, D., Kraut, G., and Wimbish, J. (1997), "The Fundamental Theorem of Statistics: Classifying Student Understanding of Basic Statistical Concepts," unpublished manuscript.
<http://www1.hollins.edu/faculty/clarkjm/papers.htm>

Cobb, G. (2005), "Forword," in *Innovations in Teaching Statistics, MAA Notes*, Vol. 65, ed. J. B. Garfield, Washington DC: Mathematical Association of America, pp. vii-viii.

Collins, L., and Mittag, K. (2005), "Effect of Calculator Technology on Student Achievement in an Introductory Statistics Course," *Statistics Education Research Journal* [Online], 4(1), 7-15.

[http://www.stat.auckland.ac.nz/~iase/serj/SERJ4\(1\)_Collins_Mittag.pdf](http://www.stat.auckland.ac.nz/~iase/serj/SERJ4(1)_Collins_Mittag.pdf)

delMas, R. C., Garfield, J., and Chance, B. L. (1999), "A Model of Classroom Research in Action: Developing Simulation Activities to Improve Students' Statistical Reasoning," *Journal of Statistics Education* [Online], 7(3). <http://www.amstat.org/publications/jse/secure/v7n3/delMas.cfm>

delMas, R., Garfield, J., Ooms, A., and Chance, B. (in press), "Assessing Students' Conceptual Understanding After a First Course in Statistics," *Statistics Education Research Journal* [Online].

delMas, R.C., and Liu, Y. (2005), "Exploring Students' Conceptions of the Standard Deviation," *Statistics Education Research Journal* [Online], 4(1), 55-82.

[http://www.stat.auckland.ac.nz/~iase/serj/SERJ4\(1\)_delMas_Liu.pdf](http://www.stat.auckland.ac.nz/~iase/serj/SERJ4(1)_delMas_Liu.pdf)

Derry, S. J., Levin, J. R., Osana, H. P., Jones, M. S., and Peterson, M. (2000), "Fostering Students' Statistical and Scientific Thinking: Lessons Learned From an Innovative College Course," *American Educational Research Journal*, 37(3), 747-773.

Dubinsky, E., and McDonald, M. (2001), "A Constructivist Theory of Learning in Undergraduate Mathematics Education Research," in *The Teaching and Learning of Mathematics at University Level: An ICMI Study*, ed. D. Holton, Dordrecht, The Netherlands: Kluwer Academic Publishers, pp. 275-282.

Fernandez, C. (2002), "Learning from Japanese Approaches to Professional Development: The Case of Lesson Study," *Journal of Teacher Education*, 53(5), 390-405.

Finney, S.J., and Schraw, G. (2003), "Self-Efficacy Beliefs in College Statistics Courses," *Contemporary Educational Psychology*, 28, 161-186.

Fong, G. T., Krantz, D. H., and Nisbett, R. E. (1986), "The Effects of Statistical Training on Thinking About Everyday Problems," *Cognitive Psychology*, 18, 253-292.

Gal, I., and Ginsburg, L. (1994), "The Role of Beliefs and Attitudes in Learning Statistics: Towards an Assessment Framework," *Journal of Statistics Education* [Online], 2(2).

<http://www.amstat.org/publications/jse/v2n2/gal.html>

Gal, I., Ginsburg, L., and Schau, C. (1997), "Monitoring Attitudes and Beliefs in Statistics Education," in *The Assessment Challenge in Statistics Education*, eds. I. Gal and J. B. Garfield, Amsterdam, The Netherlands: The International Statistical Institute, pp. 37-51.

<http://www.stat.auckland.ac.nz/~iase/publications/assessbk/chapter04.pdf>

Gall M. D., Borg W. R., and Gall J. P. (1996), *Educational Research: An Introduction*, New York: Longman Publishers.

Garfield, J. B. (1995), "How Students Learn Statistics," *International Statistical Review*, 63, 25-34.

Garfield, J. B. (1998a), "Challenges in Assessing Statistical Reasoning," Paper presented at the meeting

of the American Educational Research Association. San Diego, CA.

Garfield, J. B. (1998b), "The Statistical Reasoning Assessment: Development and Validation of a Research Tool," in *Proceedings of the Fifth International Conference on Teaching Statistics*, eds. L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee, and W. K. Wong, Singapore: International Statistical Institute, pp. 781-786. <http://www.stat.auckland.ac.nz/~iase/publications/2/Topic6u.pdf>

Garfield, J. B. (2003), "Assessing Statistical Reasoning," *Statistics Education Research Journal* [Online], 2(1), 22-38. [http://www.stat.auckland.ac.nz/~iase/serj/SERJ2\(1\).pdf](http://www.stat.auckland.ac.nz/~iase/serj/SERJ2(1).pdf)

Garfield, J. B. (2006), "Collaboration in Statistics Education Research: Stories, Reflections, and Lessons Learned," in *Proceedings of the Seventh International Conference on Teaching Statistics*, eds. A. Rossman and B. Chance, Salvador, Bahia, Brazil: International Statistical Institute, pp. 1-11. http://www.stat.auckland.ac.nz/~iase/publications/17/PL2_GARF.pdf

Garfield, J. B., and Ahlgren, A. (1988), "Difficulties in Learning Basic Concepts in Probability and Statistics: Implications for Research," *Journal for Research in Mathematics Education*, 19(1), 44-63.

Garfield, J., and Ben-Zvi, D. (in press), *Developing Students' Statistical Reasoning: Connecting Research and Teaching Practice*, Dordrecht, The Netherlands: Springer Publishing.

Garfield, J., and Chance, B. (2000), "Assessment in Statistics Education: Issues and Challenges," *Mathematics Thinking and Learning*, 2, 99-125.

Garfield, J., and delMas, R. (1994), "Students' Informal and Formal Understanding of Statistical Power," Paper presented at the Fourth International Conference on Teaching Statistics, Marrakech, Morocco.

Garfield, J., delMas, R. C., and Chance, B. (2007), "Using Students' Informal Notions of Variability to Develop an Understanding of Formal Measures of Variability," in *Thinking with Data*, eds. M. Lovett and P. Shah, Mahwah, NJ: Lawrence Erlbaum Associates, pp. 117-148.

Garfield, J., delMas, R. C., and Chance, B. (n.d.), *Assessment Resource Tools for Improving Statistical Thinking*. <https://app.gen.umn.edu/artist/index.html>.

Gigerenzer, G. (1996), "On Narrow Norms and Vague Heuristics: A Reply to Kahneman and Tversky," *Psychological Review*, 103(3), 592-596.

Giraud, G. (1997), "Cooperative Learning and Statistics Instruction," *Journal of Statistics Education* [Online], 5(3). <http://www.amstat.org/publications/jse/v5n3/giraud.html>

Gordon, S. (1995), "A Theoretical Approach to Understanding Learners of Statistics," *Journal of Statistics Education* [Online], 3(3). <http://www.amstat.org/publications/jse/v3n3/gordon.html>

Groth, R., and Bergner, J. (2005), "Pre-Service Elementary School Teachers' Metaphors for the Concept of Statistical Sample," *Statistics Education Research Journal* [Online], 4(2), 27-42. [http://www.stat.auckland.ac.nz/~iase/serj/SERJ4\(2\)_groth_bergner.pdf](http://www.stat.auckland.ac.nz/~iase/serj/SERJ4(2)_groth_bergner.pdf)

Groth, R., and Bergner, J. (2006), "Preservice Elementary Teachers' Conceptual and Procedural Knowledge of Mean, Median, and Mode," *Mathematical Thinking and Learning*, 8(1), 37-63.

Gullickson, A. R., and Ellwein, M. C. (1985), "Teacher-Made Tests: The Goodness-of-Fit Between Prescription and Practice," *Educational Measurement: Issues and Practice*, 4(1), 15-18.

Hembree, R. (1988), "Correlates, Causes, Effects, and Treatment of Test Anxiety," *Review of Educational Research*, 58(1), 47-77.

Hirsch, L. S., and O'Donnell, A. M. (2001), "Representativeness in Statistical Reasoning: Identifying and Assessing Misconceptions," *Journal of Statistics Education* [Online], 9(2).
<http://www.amstat.org/publications/jse/v9n2/hirsch.html>

Kahneman, D., Slovic, P., and Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge: Cambridge University Press.

Keeler, C. M., and Steinhorst, R. K. (1995), "Using Small Groups to Promote Active Learning in the Introductory Statistics Course: A Report from the Field," *Journal of Statistics Education* [Online], 3(2).
<http://www.amstat.org/publications/jse/v3n2/keeler.html>

Konold, C. (1989), "Informal Conceptions of Probability," *Cognition and Instruction*, 6(1), 59-98.

Konold, C. (1995), "Issues in Assessing Conceptual Understanding in Probability and Statistics," *Journal of Statistics Education* [Online], 3(1). <http://www.amstat.org/publications/jse/v3n1/konold.html>

Konold, C., Pollatsek, A., Well, A. D., Lohmeier, J., and Lipson, A. (1993), "Inconsistencies in Students' Reasoning about Probability," *Journal for Research in Mathematics Education*, 24, 392-414.

Lane, D. M., and Tang, Z. (2000), "Effectiveness of Simulation Training on Transfer of Statistical Concepts," *Journal of Educational Computing Research*, 22(4), 383-396.

Leont'ev (Leontyev) A. N. (1981). *Problems of the development of the mind*. Moscow: Progress Publishers.

Lissitz, R. W., and Schafer, W. D. (2002), *Assessment in Educational Reform: Both Means and Ends*, Boston, MA: Allyn and Bacon.

Liu, H. J. (1998), *A Cross-Cultural Study of Sex Differences in Statistical Reasoning for College Students in Taiwan and the United States*, unpublished Ph.D. dissertation.

Lovett, M. (2001), "A Collaborative Convergence on Studying Reasoning Processes: A Case Study in Statistics," in *Cognition and Instruction: 25 Years of Progress*, eds. S. Carver, and D. Klahr, Mahwah, NJ: Erlbaum.

Lunsford, M. L., Rowell, G. H., and Goodson-Espy, T. J. (2006), "Classroom Research: Assessment of Student Understanding of Sampling Distributions of Means and the Central Limit Theorem in Post-Calculus Probability and Statistics Courses," *Journal of Statistics Education* [Online], 14(3).
<http://www.amstat.org/publications/jse/v14n3/lunsford.html>

Lutzer, D. J., Maxwell, J. W., and Rodi, S. B. (2000), *Statistical Abstract of Undergraduate Programs in the Mathematical Sciences in the United States: Fall 2000 CBMS Survey*, Washington DC: American Mathematical Society.

Magel, R. C. (1998), "Using Cooperative Learning in a Large Introductory Statistics Class," *Journal of Statistics Education* [Online], 6(3). <http://www.amstat.org/publications/jse/v6n3/magel.html>

Manen, Van, M. (1990), *Researching Lived Experience: Human Science for an Action Sensitive Pedagogy*, New York: State University of New York Press.

Mathews, D., and Clark, J. (2003), "Successful Students' Conceptions of Mean, Standard Deviation and the Central Limit Theorem," unpublished manuscript. <http://www1.hollins.edu/faculty/clarkjm/papers.htm>

Meletiou, M., and Lee, C. (2003), "Studying the Evolution of Students' Conceptions Of Variation Using the Transformative and Conjecture-Driven Research Design," in *Reasoning About Variability: A Collection of Current Research Studies*, ed. C. Lee, Mt. Pleasant, MI: Central Michigan University. http://www.cst.cmich.edu/users/lee1c/SRTL3/SRLT_3_papers/SRTL-Meletiou and Lee-10-2003.pdf

Metz, K. E. (1998), "Emergent Understanding and Attribution of Randomness: Comparative Analysis of the Reasoning of Primary Grade Children and Undergraduates," *Cognition and Instruction*, 16(3), 285-365.

Meyer, O., and Lovett, M. (2002), "Implementing a Computerized Tutor in a Statistical Reasoning Course: Getting the Big Picture," in *Proceedings of the Sixth International Conference on Teaching of Statistics*, ed. B. Phillips, Voorburg, The Netherlands: International Statistical Institute. <http://www.psy.cmu.edu/LAPS/pubs/Meyer02.pdf>

Moore, D. (1998). "Statistics Among the Liberal Arts," *Journal of the American Statistical Association*, 93, 1253-1259.

Mosteller, F., and Boruch, R. (2002), *Evidence Matters: Randomized Trials in Education Research*, Harrisonburg, VA: R. R. Donnelley and Sons.

National Research Council. (2001), *Knowing What Students Know: The Science and Design of Educational Assessment*, Washington, DC: National Academy of Sciences.

Nisbett, R. E., Fong, G. T., Lehman, D. R., and Cheng, P. W. (1987), "Teaching Reasoning," *Science*, 238, 625-631.

Noss, R., Pozzi, S., and Hoyles, C. (1999), "Touching Epistemologies: Meanings of Average and Variation in Nursing Practice," *Educational Studies in Mathematics*, 40, 25-51.

Onwuegbuzie, A., and Wilson, V. (2003), "Statistics Anxiety: Nature, Etiology, Antecedents, Effects, and Treatments--A Comprehensive Review of the Literature," *Teaching in Higher Education*, 8(2), 195-209.

Pedhazur, E. J., and Schmelkin, L. P. (1991), *Measurement, Design, and Analysis: An Integrated Approach*, Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Perney, J., and Ravid, R. (1991), *The Relationship Between Attitudes Towards Statistics, Math Self-Concept, Test Anxiety and Graduate Students' Achievement in an Introductory Statistics Course*, unpublished manuscript. <http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED318607>

Pollatsek, A., Lima, S., and Well, A. (1981), "Concept or Computation: Students' Misconceptions of the Mean," *Educational Studies in Mathematics*, 12, 191-204.

Quilici, J. L., and Mayer, R. E. (2002), Teaching Students to Recognize Structural Similarities Between Statistics Word Problems," *Applied Cognitive Psychology*, 16, 325-342.

Raudenbush, S. W. (2005), "Learning from Attempts to Improve Schooling: The Contribution of Methodological Diversity," *Educational Researcher*, 34(5), 25-31.

Reid, A., and Petocz, P. (2002), "Students' Conceptions of Statistics: A Phenomenographic Study," *Journal of Statistics Education* [Online], 10(2). <http://www.amstat.org/publications/jse/v10n2/reid.html>

Reys, R.E., Lindquist, M. N., Lambdin, D. V., and Smith, N. L. (2006), *Helping Children Learn Mathematics* (8th ed.), Hoboken, NJ: John Wiley and Sons.

Roberts, D. M., and Bilderback, E. W. (1980), Reliability and Validity of a Statistics Attitude Survey, *Educational and Psychological Measurement*, 40, 235-238.

Roseth, C. J., Johnson, D. W., and Johnson, R. T. (in press), "Promoting Early Adolescents' Achievement and Peer Relationships: The Effects of Cooperative, Competitive, and Individualistic Goal Structures," *Psychological Bulletin*.

Scheaffer, R., Smith, W. B. (2007). *Using statistics effectively in mathematics education research: A report from a series of workshops organized by the American Statistical Association with funding from the National Science Foundation*. Alexandria, VA: American Statistical Association. http://www.amstat.org/research_grants/pdfs/SMERReport.pdf

Schau, C., and Mattern, N. (1997), "Use of Map Techniques in Teaching Statistics Courses," *The American Statistician*, 51(2), 171-175.

Schau, C., Stevens, J., Dauphinee, T. L., and Del Vecchio, A. (1995). "The Development and Validation of the Survey of Attitudes Toward Statistics," *Educational and Psychological Measurement*, 55, 868-875.

Sedlmeier, P., and Gigerenzer, G. (1997), "Intuitions About Sample Size: The Empirical Law of Large Numbers," *Journal of Behavioral Decision Making*, 10, 33-51.

Shaughnessy, J. M. (1992), "Research in Probability and Statistics: Reflections and Directions," in *Handbook of Research on Mathematics Teaching and Learning*, ed. D. Grouws, New York: McMillan, pp. 465-494.

Sides, A., Osherson, D., Bonini, N., and Viale, R. (2002), "On the Reality of the Conjunction Fallacy," *Memory and Cognition*, 30, 191-198.

Snee, R. (1990), "Statistical Thinking and Its Contribution to Quality," *The American Statistician*, 44, 116-121.

Snee, R. (1999), "Discussion: Development and Use of Statistical Thinking: A New Era," *International Statistical Review*, 67(3), 255-258.

Tempelaar, D. T., Gijsselaers, W. J., and Schim van der Loeff, S. (2006), "Puzzles in Statistical

Reasoning," *Journal of Statistics Education*, 14(1), 1-26.
<http://www.amstat.org/publications/jse/v14n1/tempelaar.html>

Thorndike, R. M., (2004), *Measurement and Evaluation in Psychology and Education* (7th Ed.). Upper Saddle River, NJ: Prentice Hall.

Thorndike, R. L., and Hagen, E. (1986), *Cognitive Abilities Test: Examiner's Manual Form 4*, Chicago, IL: Riverside.

Utts, J., Sommer, B., Acredolo, C., Maher, M. W., and Matthews, H. R. (2003), "A Study Comparing Traditional and Hybrid Internet-Based Instruction in Introductory Statistics Classes," *Journal of Statistics Education* [Online], 11(3). <http://www.amstat.org/publications/jse/v11n3/utts.html>

Ward, B. (2004), "The Best of Both Worlds: A Hybrid Statistics Course," *Journal of Statistics Education* [Online], 12(3). <http://www.amstat.org/publications/jse/v12n3/ward.html>

Watson, J. M., and Moritz, J. B. (2000), "Development of Understanding of Sampling for Statistical Literacy," *Journal of Mathematical Behavior*, 19, 109-136.

Weller, L. D. Jr. (2001), "Building Validity and Reliability into Classroom Tests," *National Association of Secondary School Principals Bulletin*, 85, 32-37.

Wild, C.J., and Pfannkuch, M. (1999), "Statistical Thinking in Empirical Enquiry," *International Statistical Review*, 67, 221-248.

Wise, S. L. (1985), "The Development and Validation of a Scale Measuring Attitudes Toward Statistics," *Educational and Psychological Measurement*, 45, 401-405.

Zieffler, A. (2006), *A Longitudinal Investigation of the Development of College Students' Reasoning About Bivariate Data During an Introductory Statistics Course*, unpublished Ph.D. dissertation, University of Minnesota.

Andrew Zieffler, PhD
University of Minnesota
167 EdSciB, 56 East River Road
Minneapolis, MN, 55455
zief0002@umn.edu

Joan B. Garfield, PhD
University of Minnesota
178 EdSciB, 56 East River Road
Minneapolis, MN, 55455
jbg@umn.edu

Shirley Alt
University of Minnesota
250 EdSciB, 56 East River Road
Minneapolis, MN, 55455

altx0001@umn.edu

Danielle Dupuis
University of Minnesota
250 EdSciB, 56 East River Road
Minneapolis, MN, 55455
dupui004@umn.edu

Kristine Holleque
University of Minnesota
250 EdSciB, 56 East River Road
Minneapolis, MN, 55455
holl0354@umn.edu

Beng Chang
University of Minnesota
250 EdSciB, 56 East River Road
Minneapolis, MN, 55455
beng@umn.edu

[Volume 16 \(2008\)](#) | [Archive](#) | [Index](#) | [Data Archive](#) | [Resources](#) | [Editorial Board](#) | [Guidelines for Authors](#) |
[Guidelines for Data Contributors](#) | [Home Page](#) | [Contact JSE](#) | [ASA Publications](#)