Publications

8-2012

The GAISE College Report: The American Statistical Association Meets Sound Pedagogy in Central Virginia

Beverly Wood
Embry-Riddle Aeronautical University, woodb14@erau.edu

Scholarly Commons Citation

The GAISE College Report: The American Statistical Association

Meets Sound Pedagogy in Central Virginia

_____

A Dissertation

Presented to

The Faculty of the Curry School of Education

University of Virginia

_____

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

_____

by

Beverly L. Wood

M.S., University of Wisconsin-Oshkosh, 2003
B.S., University of Tampa, 1988

August 2012

UMI Number: 3528782

![UMI Dissertation Publishing logo]

UMI 3528782

![ProQuest logo]

Abstract

Advisor: Robert Q. Berry, III, Ph.D.

Research in undergraduate statistics education often centers on the introductory

course required for a large percentage of college students. While acknowledging the

diverse setting, audience, and purpose of introductory courses, existing research assumes

that courses offered by different disciplines share the same goals and teaching practices.

The purpose of this study is to examine the objectives for student outcomes and

pedagogical delivery of introductory statistics courses in various academic departments to

provide explicit evidence for this assumption.

The American Statistical Association's *Guidelines for Assessment and Instruction

in Statistics Education* (GAISE) are meant to apply to all introductory courses. The

College Report's Goals for Students and Recommendations for Teaching are used as a

framework for a qualitative study of the way in which introductory courses in various

settings deliver instruction. Four descriptive case studies are presented through a pattern-

matching analysis followed by a cross-case analysis.
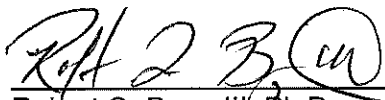
All four cases demonstrate many of the goals and teaching strategies

recommended by GAISE, even though none of the professors had prior knowledge of the

guidelines. The goal that students be able to critique published statistics resonated with

participating instructors but was barely evident in any of the courses. The

recommendation to use real data had the least evidence in all cases. Emphasis on

statistical literacy and thinking as well as stress on conceptual understanding aligned with

GAISE in every case. This study supports the GAISE assumption that its goals for

students and recommendations for teaching are broad enough to apply to introductory

courses in a variety of disciplines.
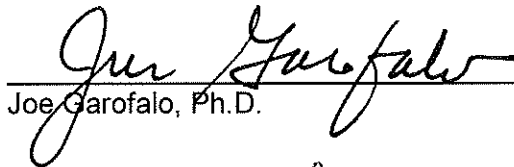
Curriculum, Instruction, and Special Education
Curry School of Education
University of Virginia
Charlottesville, Virginia

APPROVAL OF THE DISSERTATION
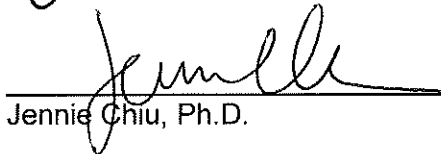
This dissertation, *The GAISE College Report: The American Statistical Association Meets Sound Pedagogy in Central Virginia,* has been approved by the Graduate Faculty of the Curry School of Education in partial fulfillment of the requirements for the degree of Doctor of Philosophy.
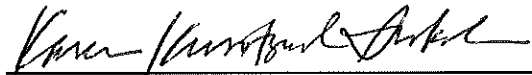
Robert Q. Berry III, Ph.D.

Joe Garofalo, Ph.D.

Jennie Chiu, Ph.D.

Karen Inkelas, Ph.D.

6/26/12 _____ Date

To the One who gives me
strength to do all things.

and

To my husband who tells me
"you're braver than you believe,
and stronger than you seem,
and smarter than you think."

Acknowledgements

While my family supported my physical needs and professors supported my intellectual needs, new friends in Charlottesville met my social needs. A group of fellow students—thrown together our first semester of coursework in spite of our disparate areas of study—formed a supportive cohort that made difficult times better and good times more fun. Thank you Kate Dabney, Natasha Heny, Bong Gee Jang, Colby Tofel-Grehl, and Peter Wiens for this unique brand of friendship. Wendi Dass has been a classmate, project partner, presentation collaborator, employment reference, and, in the end, peer reviewer extraordinaire. Many other Curry students contributed to my academic survival in shorter, but no less important, times of need. The people of Northridge Community Church offered a warm welcome and friendly ear on many occasions.

Finally, there would be no analysis to perform if not for the generosity of four busy professors teaching statistics to undergraduates. Their candor about teaching and interest in my research made the data collection phase far easier than it might have been. I hope that they feel satisfaction in the inherent self-reflection they engaged in during my probing into their classes.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction

Statement of the Problem

Adult literacy is an important pillar of democracy. Thomas Jefferson wrote to James Madison in 1787 that an informed citizenry is "the only sure reliance for the preservation of liberty" (as cited in Steen, 1997). Though the sentiment remains intact, the description of an informed citizenry has changed dramatically in the two centuries following Jefferson's assertion. The ever-changing social and economic environment in which citizens must function necessitates a constant revision of what it takes to be literate or informed.

The Young Adult Literacy survey (YALS) of 1985 set the current standard for literacy assessment by reporting the results in terms of three scales: prose, document and quantitative (Campbell, Kirsch, & Kolstad, 1992; Shaughnessy, 2007; Steen, 2004). The first scale measures the knowledge and skills needed to glean information from a variety of textual sources. Knowledge and skills required to locate and use information presented non-textually (tables, graphs, maps, forms) are identified on the document scale. The quantitative scale applies to the knowledge and skills necessary to apply arithmetic operations to numbers embedded in text (Campbell, et al., 1992; Kirsch & Jungblut, 1986; Kirsch, et al., 1993). In reporting the results of this large-scale study, adult literacy was no longer viewed as a single construct but as three intricately-

connected yet separately-measureable literacies (Kirsh & Jungblut, 1986; Kirsch, et al., 1993).

An expanded study across all adult age groups (the National Adult Literacy Survey or NALS) was conducted in 1992. This survey was designed to allow direct comparisons with YALS for the purpose of identifying improvement (Kirsch, et al., 1993). Both surveys found low levels of both document and quantitative literacy—about 50% of adults performing at Intermediate or Proficient levels—although more than 90% of participants could read short, simple text that would have categorized them as "literate" by earlier standards (Kirsh & Jungblut, 1986; Kirsch, et al., 1993).

An introductory statistics course contains elements of all three types of literacy. Prose literacy is required to take in new information about statistical concepts and procedures through the textbook and/or lecture notes, as well as to understand scenarios that require statistical analysis. Document literacy is incorporated into an introductory course both as sources of data (tables, arrays) and as communication of information generated by data (graphs, charts). The introductory course demands quantitative literacy in order to implement statistical procedures for making sense of data and in order to make decisions based on it. Statistics courses, therefore, are positioned to make an impact on Adult Literacy measures through the practice and application of all three scales.

The American Statistical Association (ASA) recognized the important role that statistics education would play in the quest for an informed citizenry (Ben-Zvi & Garfield, 2008). The 1980s saw the ASA cooperating with the National Council of Teachers of Mathematics (NCTM) in an effort to infuse data analysis and rudimentary statistics into school curricula. This cooperative effort was called "The Quantitative

Literacy Project" (Scheaffer, 2003; Steen, 2001). The Mathematical Association of America (MAA) also expressed interest through its Curriculum Action Project and George Cobb's email focus group on statistics education (American Statistical Association, 2005; Cobb, 1992; Scheaffer, 2003). George Cobb recommended changes for college-level introductory statistics courses in the face of increasing access to computing equipment as well as changes in professional practice and theory (Cobb, 1992).

In 2003, the ASA funded the Guidelines for Assessment and Instruction in Statistics Education (GAISE) Project to develop a set of guidelines for the introductory statistics course. The College Report prefaces the list of goals for students with an overarching vision: "The desired result of *all* [emphasis added] introductory statistics courses is to produce statistically educated students, which means that students should develop statistical literacy and the ability to think statistically" (American Statistical Association, 2005, p. 11). The goals for students and recommendations for teaching introductory courses in statistics acknowledge the reality that statistics is "a family of courses, taught to students at many levels, from pre-high school to post-baccalaureate, with very diverse interests and goals" (ASA, 2005, p. 7) and, therefore, does not present a list of topics to be covered but general principles for focusing *any* course on the statistical literacy and thinking of its students.

There are two tacit assumptions in much of the research on statistics education regarding the introductory statistics course: 1) the objectives of introductory statistics courses are primarily focused on students' general education; and 2) the academic context of course offerings is a non-salient feature to the acquisition of statistical literacy and

3

ability to think statistically. The GAISE College Report (ASA, 2005) describes some of the diversity in pedagogical style and emphasis on statistical literacy that can be found in introductory courses and suggests that the goals and recommendation endorsed by the ASA apply to them all.

Much of the diversity in introductory courses is a consequence of the diverse history of the development of statistics as a discipline in its own right. Many statistical tools and techniques were introduced by professionals in fields such as biology (e.g., correlation coefficient), chemistry (e.g., Student's t distribution), agriculture (e.g., ANOVA), and economics (e.g., multiple collinearity). The documented evolution of statistics courses at Oklahoma State University illustrates this diversity. In the 1926-27 academic year a course entitled "Biometry" was offered by the Department of Field Crops and Soils for the first time. The three succeeding years added "Business Statistics" to the Business Administration curriculum, "School Statistics" as a graduate course for Education Administration, and "Theory of Least Squares," also as a graduate course, but in the Mathematics Department (Folks, 2002).

The diaspora of introductory statistics course offerings can be found at institutions of all sizes. The University of Virginia—a large, research-intensive institution—offers nine courses at the undergraduate level. My *alma mater*, the University of Tampa, is now a medium-sized master's institution that offers seven undergraduate courses in statistics. A small institution in Virginia, Marymount University, offers four courses to its undergraduates, and even Piedmont Virginia Community College provides three options.

The 2005 report from the Conference Board of the Mathematical Sciences (Lutzer, Rodi, Kirkman, & Maxwell, 2007) confirms a 9% increase in enrollments in

elementary-level (non-calculus) statistics courses at four-year institutions and a 58%

increase at two-year institutions since its 1995 report. Like the ones that preceded it, the

2005 report only deals with courses offered by Mathematics and/or Statistics

Departments. There is not a comprehensive report available to identify parallel increases

in enrollment for introductory statistics courses offered by other disciplines; however, in

his chapter on statistics in the *Second Handbook of Research on Mathematics Teaching

and Learning*, J. Michael Shaughnessy (2007) claims that "statistics is required in almost

all collegiate majors" (p. 1000).

Even if Shaughnessy's claim exaggerates the proportion of students required to

take statistics, the fact that the total undergraduate student enrollment in degree-granting

institutions that year was over 18 million (National Center for Education Statistics, 2009)

indicates that the quality of introductory statistics education impacts millions of

undergraduates. The NCES report also notes that in 2006 although only "12 percent of

the campuses enrolled 10,000 or more students, they accounted for 55 percent of total

college enrollments" (2009, p. 270). A study of statistics courses at a large university

may shed light on statistics education opportunities for a large portion of those millions.

Attention to courses offered by smaller institutions expose additional nuances in how

introductory courses vary due to their academic environment.

Purpose

The purpose of this project is to delve into objectives for student outcomes and

pedagogical delivery of introductory statistics courses in various academic departments

through multiple case studies. Comparisons across the cases inform the validity of the

research assumptions previously noted in light of the distributed structure of statistics

education at the tertiary level. The GAISE College Report offers a framework that rests on those same assumptions to assess individual course alignment to the Goals for Students (see page 3 of the Observation Protocol, Appendix A, for a complete listing) and Recommendations for Teaching (page 1 of Appendix A).

Research Questions

Two questions emerge from the intersection of the growth of statistics education research and the publication of Guidelines from the ASA:

- How do the introductory statistics courses offered by different academic departments define objectives and deliver instruction?

- Are there sufficient commonalities for students in *all* classes to achieve the level of statistical literacy and thinking recommended by the GAISE College Report?

Significance of the Study

Answering these questions provides evidence for the validity of the assumptions made by both statistics education researchers and the American Statistical Association regarding "the introductory course" in its diverse settings and with its diverse content. Findings that do not support the assumptions provide new information for continued discussion about the "who", "what", "where", and "how" of undergraduate statistics education. Findings that do support the assumptions offer illustrations of the diversity within the family of courses that GAISE addresses.

Operational Definition of Terms

There are a variety of terms used to talk about a beginning course in statistics, though the word "beginning" is rarely among them. Statisticians, statistics educators, and statistics education researchers may have nuanced ideas about three of the terms that will

be used in the discussion of this study. To be clear about their usage, operational definitions for this report are given below.

Introductory course: A course designed to provide thorough coverage of descriptive statistics, some probability topics, and the basics of inferential statistics, usually in the form of confidence intervals and hypothesis testing. There may or may not be a prerequisite of calculus.

Elementary course: A course designed to provide thorough coverage of descriptive statistics, very little probability, and a rudimentary treatment of inferential topics. No calculus is expected of the students; may also be referred to as algebra-based.

First course: May resemble the introductory or elementary course but with the added assumption that the students have not had any previous formal exposure to the course content and will continue on to at least a second course.

A fourth variety is more difficult to define. A **Data Analysis course** may refer to a course that focuses on descriptive statistics and exploratory data analysis techniques. The term might also indicate a broader course that relies on computer analysis, which may have a prerequisite statistics course. In order to avoid that ambiguity, the term will not be used in this study.

The goals and recommendations of the GAISE College Report (ASA, 2005) uses the term *introductory* to mean all courses without another statistics course as a prerequisite, including those offered in high school and graduate or professional schools. This usage defines courses eligible for inclusion in this study but, once included, the courses will be distinguished as defined above.

Three other terms with diverse definitions in the research literature need to be clarified at the outset of this study:

Quantitative literacy refers to an individual's ability to glean numerical information from a variety of sources and apply that information to decision-making situations in their personal lives (e.g., finances, transportation), within their employment situation (e.g., accounting, personnel management), and as informed citizens (e.g., voting, political debate).

Statistical literacy will be considered as a subset of quantitative literacy, referring specifically to numerical information that results from statistical procedures (e.g. interval estimates, risk analysis).

Statistical thinking goes beyond the use of statistical results to the practice of considering statistical analysis useful for informing problems involving data and uncertainty.

Approaching the Literature

The Guidelines for Assessment and Instruction in Statistics Education did not arise in a vacuum. A look at its predecessors, the growth of statistics as a scientific tool, and the necessity for adequate preparation of literate citizens is necessary. The following chapter will delve into these areas.

Chapter 2: Review of the Literature

The purpose of this chapter is to provide a thorough description of the conceptual context for this study and to investigate research findings related to it. One stream of research is the evolving need for citizens to deal with statistical ideas in their daily lives. Researchers in this stream are concerned about school mathematics, adult education, workforce development, and remedial college mathematics. Another stream is the emergence of statistical practice out of multiple disciplines. Researchers in this stream include mathematical statisticians, applied statisticians, and statistics educators within colleges and universities. At the confluence of the two research streams is the reformation of statistics education. Researchers find that their interests converge here because distinctions between quantitative literacy in adults and statistical education of tertiary students are muddy. The American Statistical Association's endorsement of the Guidelines of Assessment and Instruction in Statistics Education is an important marker in the flow of statistical sophistication expected of university graduates in both their professional and personal lives.

Figure 1. Interest convergence. This figure illustrates the
convergence of two academic pursuits and their mingling
within the introductory statistics course.

Adult Literacy, Quantitative Literacy, and Statistics Education

It has already been noted that our founding fathers deemed education an important

pillar of democracy. Two centuries have not changed the need for an informed citizenry,

but have transformed the notion of what constitutes an educated, and therefore informed,

citizen. The National Governors' Association met in 1990 to establish a set of National

Education Goals and took note of the founders' concern:

> By the year 2000, every adult American will be literate and will
> possess the knowledge and skills necessary to compete in a global
> economy and exercise the rights and responsibilities of citizenship.
> (Campbell, et al., 1992)

Defining what it means for a person to "be literate" or for a nation to possess an

"informed citizenry" is a difficult task. The ever-changing social and economic

environment in which citizens must function necessitates a constant revision of what it

takes to be literate or informed. As the United States evolved from farming to

commercialism to industrialism to a knowledge-based economy, the literacy required of its citizens grew in complexity and the level of state-sponsored education grew in response. Though hardly universal, literacy has been sufficiently widespread to sustain the nation and to outclass the rest of the world (Ellis, 2001).

How to Measure Adult Literacy

Measuring literacy has likewise grown in complexity. Historians have used counts of signatures on wills, marriage licenses and deeds to estimate early literacy rates. The U.S. Census Bureau began collecting self-reported literacy information in the mid-1800s. Standardized tests of school-based reading skills took hold after the entrance tests for Army recruits in World War I belied the self-reported rates from the Census. In the 1970s, competency-based surveys finally included measures of computation, problem solving, and interpersonal skills to gauge more accurately the ability to meet challenges that adults typically encounter at home, at work, or in the community (Campbell et al., 1992).

The Young Adult Literacy survey of 1985 set the current standard for literacy assessment by reporting the results in terms of three scales: prose, document and quantitative (Campbell et al., 1992; Shaughnessy, 2007; Steen, 2004; Tolbert-Bynum, 2008). The first scale measures the knowledge and skills needed to glean information from a variety of textual sources. Document literacy identifies the knowledge and skills required to locate and use information presented non-textually (tables, graphs, maps, forms). The quantitative scale applies to the knowledge and skills necessary to apply arithmetic operations to numbers embedded in text (Campbell et al., 1992).

An expanded study across all age groups (the National Adult Literacy Survey or NALS) was conducted in 1992. A similar international study (International Survey of Adult Literacy, ISAL) was taken at about the same time. Both studies, like the Young Adult Literacy Survey that preceded them, revealed low levels of both document and quantitative literacy among U.S. adults. The assessments subdivided the tasks into five levels of difficulty and found discouragingly small percentages of Americans performing at the top two levels (Dossey, 1997).

NALS was purposely constructed so that direct comparisons could be made with YALS for the purpose of identifying improvement (Kirsch, et al., 1993). Of particular concern was the *decrease* in average scores on all three scales from the 1985 survey to the 1992. This was evident in a comparison of the 21-25-age group of the two surveys as well as in a comparison of the 1985 21-25 age group to the 1992 28-32 age group that represented the same cohort (Kirsch, et al., 1993).

How to Improve Adult Literacy

Major professional organizations interested in mathematics education responded to the dismal results with conferences, forums and published works in the late 1980s and throughout the 1990s. The Mathematical Association of America, National Council on Education and the Disciplines, The College Board, the American Statistical Association, and National Council of Teachers of Mathematics continue to support calls for recognition that quantitative literacy is as important to effective citizenship – as well as an economic advantage to the individual citizen – as prose literacy. Varying definitions of what exactly comprises quantitative literacy does not impede a unity as to its importance or the need for its development outside the mathematics classroom

(Bookman, Ganter, & Morgan, 2008; Burke, 2007; Madison, 2004; McClure & Sircar, 2008; Steen, 1997, 2001, 2004; Wiest, Higgins, & Frost, 2007). Wade Ellis captures the common thread among these powerful organizations and numerous researchers: "For me, quantitative literacy is more like art than science. I know it when I see it, but I cannot easily define it" (2001).

Current Practices in Quantitative Literacy (Gillman, 2006) provides eleven essays on how quantitative reasoning is infused interdepartmentally at individual institutions, as well as seven essays that address a specific course available to an institution's undergraduates. This text is published by the Mathematical Association of America and is unsurprisingly heavy on courses taught by or in collaboration with mathematics departments. Another not-unexpected theme is the inclusion of topics related to probability, exploratory data analysis, and critical awareness of statistical claims.

Statistical literacy, like quantitative literacy, is not a precisely defined term in the extant literature. Shaughnessy (2007) does note the agreement of researchers that the ability to respond to statistical information and to critique it is a hallmark of this type of literacy. The *Second Handbook* provided a quote from Watson and Moritz: "Judging statistical claims from the media is fundamental to being statistically literate" (as cited in Shaughnessy, 2007). Another widely accepted characteristic of quantitative literacy is that most of the work is middle school mathematics (Steen, 1997, 2001, 2006; Wiest et al, 2007). Many students succeed in memorizing statistical procedures that require only basic mathematics but few understand the work they are doing and how it is evident in their everyday lives.

Researchers in statistics education have investigated the quantitative literacy implications of their courses (Gal, 2002). Some offer evidence of effective pedagogy to enhance statistical literacy (Chiou, 2009; Meyer & Dwyer, 2005; Root, 2009), while others investigate factors influencing student acquisition of statistical literacy (Gnaldi, 2006; McClure & Sircar, 2008; Wade & Goodfellow, 2009). Shaughnessy (2007) reviewed recent research on statistical learning and reasoning in his chapter of the *Second Handbook of Research on Mathematics Teaching and Learning*, including a wide array of studies dealing with "statistical literacy" at both the secondary and tertiary levels.

Historical Background of University Statistics Education

Slow beginnings

Probability and statistics are arguably as old as human society. Games of chance date back to at least 3000 B.C. but probability received no scholarly attention until the 16th century A.D. (David, 1970). Societies have been counting people—for tax collection and army-raising purposes—for nearly as long (e.g., Exodus 30:12, New International Version of the Bible), again without scholarly attention until the 17th century A.D. when work in demographic and actuarial sciences began (Heyde & Seneta, 2001). "Many eighteenth-century scientists had at least a vague feeling that probability would underlie an eventual successful treatment of social data… but as a tool for the reduction and measurement of uncertainty in data, the calculus of probability had proved largely sterile" (Stigler, 1986, p. 99) until the dawning of the 19th century brought a general central limit theorem (Heyde & Seneta, 2001) on which to build the desired bridge from descriptive to inferential statistics.

The late 1700s brought the first publication of graphs and charts as summaries of data (Spence & Wainer, 1997). By the mid-1800s the floodgates were opened to the use of carefully collected, well organized, and clearly summarized statistics as a vehicle for social change. Florence Nightingale combined her unusual (for a woman of the age) mathematics training, her compassion as a nurse, and her family connections to influence Queen Victoria to commission change in the hospital conditions of the army (Heyde & Seneta, 2001; O'Connor & Robertson, n.d.). The use of statistics to effect change had arrived!

Arrival of Statistics at the University

Mathematicians both in (e.g., Bernoulli, Chebyshev, Poisson) and out (e.g., Bayes, DeMoive, Fermat) of the university made great contributions to the development of probability theory. It was, however, a much broader variety of scientists making contributions to the evolution of statistics in the 19th and early 20th centuries. Most of them had some mathematics training that they wished to apply to data collected in their primary discipline.

Karl Pearson is the lynchpin for turning statistics from a mathematical curiosity to a subject of study in its own right. His initial degree in mathematics from Cambridge was followed by further studies in philosophy, physics, metaphysics, law, and German. After passing the bar, he briefly practiced law then lectured on German for a couple of years. In the spring of 1884 he was offered a post in German at Cambridge, which he declined, preferring the Chair of Mechanisms and Applied Mathematics at University College London (Heyde & Seneta, 2001).

From 1891 to 1893 Pearson also held the Gresham Chair of Geometry, which required twelve public lectures a year. It is in these Gresham Lectures that he collected and presented statistical procedures that are still common in today's introductory courses. The procedures themselves were not new but some of the vocabulary was, for example: histogram (a time diagram to be used for historical purposes), standard deviation (rather than mean error), and normal curve (instead of curve of error) (Heyde & Seneta, 2001; Pearson, 1936). These lectures introduced Pearson to Raphael Weldon and Francis Galton who were interested in statistical methods for their own work in evolutionary biology and ancestral heredity, respectively (O'Connor & Robertson, n.d.).

Pearson founded the Biometric School in 1892 where modern statistics was incubated. This evolved into the Biometric Laboratory where brewery chemist William Sealy Gosset (the famous Student with a t distribution) came to study in 1908. The long-held conviction that biological measurements followed the distribution of the normal curve was challenged by Pearson's prolific presentation of empirical evidence of distributions that are J- or U-shaped or definitely skewed from normal. Publishing nearly 400 papers on statistics, Pearson offered a plethora of methods that are still in use today: simple regression, standard error of an estimate, correlation coefficient, multiple and partial correlation, multiple regression, biserial correlations, and χ^2 tests (Heyde & Seneta, 2001; O'Connor & Robertson, n.d.).

From Pearson's Biometric School and Laboratory, statistics became a part of the university curriculum. In 1911 University College London founded a Department of Applied Statistics with Pearson as its head and by 1915 the first degree in statistics was offered (Department of Statistical Science, 2008). Following the establishment of

scholarly journals by the Royal Statistical Society and the American Statistical Association, which had begun in the 19th century (Royal Statistical Society, n.d.; Mason, 1999), the degree course settled statistics into the life of the university.

George Snedecor founded the first university unit dedicated to statistics in the United States at Iowa State College (now, University) in 1927. The Mathematical Statistical Service grew into the Statistical Laboratory in 1933 and in 1947 became the Department of Statistics (Heyde & Seneta, 2001; Hobbs, 2008). Snedecor began teaching the first course completely dedicated to statistics, "Mathematical Theory of Statistics," in 1914-15 when he was promoted to Associate Professor of Mathematics in his second year at the University. The course evolved into two before the end of the decade and by the early 1920s other departments on campus were offering their own courses (Heyde & Seneta, 2001).

The first M.S. degree in Statistics was awarded by the Mathematics Department of Iowa State College to Gertrude Cox in 1931. Her thesis was titled *A Statistical Investigation of a Teacher's Ability as Indicated by the Success of His Students in Subsequent Courses* (Heyde & Seneta, 2001; O'Connor & Robertson, n.d.). She worked for the Statistical Laboratory until 1940 when she became the first woman professor at North Carolina State College (now, University) and founder of its Department of Statistics the following year (Department of Statistics, n.d.; Heyde & Seneta, 2001; Hobbs, 2008; O'Connor & Robertson, n.d.).

Both Snedecor and Cox wrote enduring textbooks for their statistics students: *Statistical Methods* in 1937 and *Experimental Design* in 1950, respectively. William G. Cochran collaborated with Snedecor then co-authored with Cox as his academic career

shifted from Iowa State to North Carolina State (Heyde & Seneta, 2001; Hobbs, 2008; O'Connor & Robertson, n.d.). Along with R.A. Fisher's 1925 classic *Statistical Methods for Research Workers* (Heyde & Seneta, 2001; O'Connor & Robertson, n.d.), statistics educators had excellent texts from which to choose through most of the 20th century for the training of future statisticians.

Reformation of Statistics Education

The American Statistical Association (ASA) established its Section on Statistics Education in 1944 (Mason, n.d.), thus recognizing the importance of post-secondary education to the development of its profession. This occurred not long after American university degrees in statistics were first offered in the 1930s, and immediately[1] following the founding of the earliest Departments of Statistics in the United States (Heyde & Seneta, 2001). At the midpoint of the twentieth century, the ASA caught H.G. Wells' vision that "Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write" (as paraphrased by Wilks, 1950) and began to consider the statistical needs of all Americans, not just professional statisticians.

The 1980s saw the ASA cooperating with the National Council of Teachers of Mathematics (NCTM) in an effort to infuse data analysis and rudimentary statistics into school curricula. This cooperative effort was called "The Quantitative Literacy Project" in response to the national interest in improvement of adult literacy as defined by three scales: prose, document, and quantitative (Scheaffer, 2003; Steen, 2001). The Mathematical Association of America (MAA) also expressed interest through its Curriculum Action Project and an email focus group on statistics education (ASA, 2005; Cobb, 1992; Scheaffer, 2003). With the increasing access to computing equipment as

---

[1] discounting the War Years when the Association's annual meeting were cancelled (Mason, n.d.)

well as changes in professional practice and theory, George Cobb recommended changes for college-level introductory statistics courses (Cobb, 1992).

Cobb's three recommendations were 1) emphasize statistical thinking; 2) present more data and concepts, less theory and fewer recipes; and 3) foster active learning (Cobb, 1992). The emphasis on statistical thinking was further detailed as instruction to help students understand the need for data, the importance of data production, the omnipresence of variability, and the quantification/explanation of variability. A survey of instructors of introductory statistics courses conducted at the end of the decade demonstrated the impact of Cobb's recommendations (Garfield, 2000). In light of the rapidly expanding enrollments in undergraduate and high school Advanced Placement statistics courses (Lutzer, et al., 2007; Shaughnessy, 2007), the ASA funded a project to produce evidence-based Guidelines for Assessment and Instruction in Statistics Education (GAISE) for primary and secondary education with a separate report for tertiary courses. The GAISE College Report built explicitly upon Cobb's recommendations[2] to produce their six recommendations:

1. Emphasize statistical literacy and develop statistical thinking.
2. Use real data.
3. Stress conceptual understanding rather than mere knowledge of procedures.
4. Foster active learning in the classroom.
5. Use technology for developing conceptual understanding and analyzing data.
6. Use assessments to improve and evaluate student learning. (ASA, 2005, p. 4)

Evidence-based pedagogy.

Readers familiar with modern educational research will find recommendations three through six to be completely consistent with current ideas about quality teaching. A recent development in mathematics education is a framework for teachers to reflect on

---

[2] George Cobb was a member of the GAISE committee.

their teaching in ways that align with NCTM's *Mathematics Teaching Today: Improving Practice, Improving Student Learning* (NCTM 2007) and *Principles and Standards ' for School Mathematics* (NCTM 2000). The nine dimensions overlap the GAISE recommendations for teaching in both obvious and subtle ways. An example will suffice to illustrate the similarity and overlap:

> Multiple representations: Are a variety of representations (graphs, pictures, symbols, charts, diagrams, or manipulatives) used during instruction?
>
> Use of mathematical tools: Do students have the opportunity to use appropriate math tools (other than paper, textbooks, or chalkboards) to investigate concepts and solve problems in class? (p. 241)

By considering statistical software and data sets as tools, the multiple ways of summarizing and presenting data as well as using simulations to investigate concepts overlaps with the GAISE recommendations three, four, and five. See Merritt, Rimm-Kaufman, Berry, Walkowiak, & McCracken (2010) for the complete list and thorough operational definitions.

Latest Directions

A simple search of three databases (Education Research Complete, ERIC, and Academic Search Complete) using *GAISE* as the only search term yielded 35 documents that referred to the ASA's Guidelines. Twelve items, including two book chapters, were focused on pre-college instruction and were set aside for later review. It is not surprising that many authors of the articles are familiar names in the statistics education research community. They have long encouraged statistics educators to emphasize statistical literacy and thinking; to develop conceptual understanding over procedural knowledge; and to use technology, authentic assessment, and real data to do so (Chance & Rossman, 2001; delMas, Garfield, & Chance, 1999; Garfield, 1995; Petocz & Reid, 2003; Rumsey, 2002; Utts, 2003; Wild &

Pfannkuch, 1999). The remaining 23 articles were easily separated into three categories:

conceptual (7), practical (9), and empirical (7).

Conceptual

Petocz and Reid (2005) initiate their call for reform in the tertiary mathematics

curriculum with their belief that there is room for enhancement "by a reorientation towards

one that treats students as citizens of the world first" (p. 89). They draw attention to the

MAA's 2004 *Curriculum Guide* and a draft of the GAISE recommendations as evidence that

their suggestions do not stand alone. The GAISE College Report particularly puts the

students at the center of instruction that equips statistically literate citizens, meeting Petocz

and Reid's vision for pedagogical enhancement within the mathematical sciences.

Hassad (2008) defines reform-oriented teaching as pedagogy that is in alignment with

the GAISE recommendations. The purpose of his article is to encourage the health, social,

and behavioral science disciplines to see the need for reform and adopt pedagogy that fosters

the statistical literacy necessitated by emerging recognition of the importance of evidence-

based practice in those disciplines. He concludes with a call for promotion and tenure

committees to see the value of curricular development as further encouragement for the

adoption of reformed pedagogy in the introductory statistics courses.

The ASA's Section on Statistical Education hosted a panel discussion at the 2006

Joint Statistical Meetings regarding student retention of important statistical ideas and how to

assess that retention (Berenson et al., 2008). Two of the panelists, Mark Berenson and Karen

Kinard, directly addressed the GAISE sixth recommendation: Use assessment to improve

and evaluate student learning. Two others, Jessica Utts and Deborah Rumsey, address

assessment through discussion of retention as a result of emphasis on conceptual

understanding and active learning, the third and fourth recommendations.

Hall and Roswell (2008) used the GAISE recommendations as a framework to evaluate the support that the National Science Foundation has provided for statistics education reform. They found 110 projects funded in the decade preceding the publication of the GAISE College Report, noting that 95% of them met at least one of the six recommendations. Attesting to the inter-connectedness of the recommendations is the fact that 65% of the projects met more than one, even when the researchers were focused on one recommendation in particular.

Joan Garfield and Robert delMas (2010) introduce their article on resources for assessment of statistical thinking with the sixth GAISE recommendation. Joan Garfield and Michelle Everson (2009) describe a unique graduate-level course for preparing teachers of statistics and its alignment with the GAISE recommendations. Michelle Everson, Andrew Zieffler and Joan Garfield (2008) discuss ways in which introductory courses can be changed in order to better reflect the ASA's vision for effective instruction. This research group from the University of Minnesota consistently blends the conceptual and the practical as evidenced in these papers.

Practical

Richardson, Stephenson, and Gabrosek (2010) describe the use of the golf-dice game GOLO as an activity to illustrate descriptive statistics, both numerical and graphical. They specifically link the game to GAISE recommendations two, three, and four (use real data, conceptual understanding, and active learning). The same group of statistics education researchers had previously described the use of the game as an illustration of Cobb's components of statistical thinking that were quoted in the GAISE report (Gabrosek, Stephenson, & Richardson, 2008). The earlier publication was geared to high school teachers but the activity was developed for a tertiary course as was the use

in descriptive statistics (Gabrosek et al., 2008; Richardson et al., 2010). These two articles indicate the similarities in general statistics courses at the high school level and the introductory courses at the college level.

Pfannkuch, et al., (2010), Delcham and Sezer (2010), and Sisto (2009) address the challenge of language during reform-oriented statistics instruction. The first two articles describe different styles of writing projects to be used as evidence of the statistical literacy of students. In both cases the writing projects were included in the authors' courses after revisions based on the GAISE recommendations. Sisto discusses the increased challenge of both verbal and written expression of statistical ideas in the context of a group project in a multicultural classroom attempting to meet GAISE recommendations. All of these activities mention the well-documented difficulties of vocabulary in statistics instruction (see Kaplan, Fisher, & Rogness, 2009 for an overview of that literature).

A pair of articles present course projects specifically designed to meet the second GAISE recommendation: Use real data. Nelson (2009) prefers to reference an expanded recommendation, "Use real data that tell a compelling story" (p. 1), which is also evident in the project described by Fink and Lunsford (2009). Both examples relate to the environment, providing current and relevant context to the statistical concepts presented via group projects, incidentally meeting GAISE recommendation four: Foster active learning in the classroom. In many ways, these two activities are complementary since one uses a large, existing data source while the other requires firsthand data collection. Students experience different types of challenge during the projects, all of which are

intended to develop statistical thinking in line with GAISE recommendation one: Emphasize statistical literacy and develop statistical thinking.

The full report from the GAISE committee prefaces the recommendations with a list of "Goals for Students in an Introductory Course: What it Means to be Statistically Educated" (p. 5). It is against this list that Allan Rossman (2009) offers four examples of activities to illustrate topics of inference. The first of these activities draws upon student intuitive understanding through the use of dishonest dice to model "Fisherian inductive reasoning" (p. 7); the other three describe the use of stochastic simulation as an alternative method of inference to the ubiquitous use of hypothesis testing. The reference list of this article contains multiple sources for better understanding of the Fisher and Neyman methods for inference, including the original publications and more recent debate among professional statisticians.

Holt and Scariano (2009) offer a mathematically sophisticated activity utilizing the probability density function for determining the "best" measure of center for a realistic situation in statistical consulting. The activity described is intended for the post-calculus student with adaptations for students above and below this level of mathematical maturity. In the introduction to the article, Holt and Scariano discuss the GAISE recommendations as applicable to courses other than the algebra-based elementary course that receives most of the attention of statistics education researchers working at the post-secondary level. Few researchers explicitly distinguish between the elementary and introductory course in statistics, often assuming that all introductory courses are algebra-based. Holt and Scariano remind the research community that this assumption is faulty. *Empirical*

Of the seven empirical articles, two reported research not related to student outcomes after GAISE-compliant teaching. Green (2010) conducted a qualitative study of the training given to Teaching Assistants (TAs) at the University of Nebraska-Lincoln where they are given full teaching responsibility for the introductory course rather than a true assistantship. Encouraged by similar work at the University of Minnesota (see Garfield & Everson, 2009) and the results of Green's findings, the University of Nebraska-Lincoln developed a course "to help TAs develop effective strategies for teaching statistical concepts aligned with the GAISE guidelines" (p. 119). Chiesi and Primi (2010) reported on a study conducted with psychology students enrolled in introductory statistics to investigate the cognitive and non-cognitive factors influencing course achievement through a structural equation model. The course, however, was not described as being designed with the GAISE recommendations in mind. Rather, their mention of GAISE was as a contrast to their suggestion to provide students with *additional* mathematics instruction as part of the introductory course.

Four of the final five articles resonate with those suggesting activities that are reviewed above. Lesser and Winsor (2009) conducted qualitative research on the experiences of English Language Learners (ELLs) in an introductory statistics course. Some of his findings address the ambiguity of statistical vocabulary consistent with Sisto (2010) and Kaplan et al. (2009). In his discussion, the GAISE recommendation for active learning and its benefit of providing practice with statistical communication may prove particularly beneficial for ELLs.

Two studies addressed course assignments related to written language. Neumann and Hood (2009) studied the use of wikis as consistent with the GAISE goals. Theoret

and Luna (2009) studied the use of journals and discussion boards with the same goals in mind. Many of the measures used by Neumann and Hood did not show a statistically significant difference between the wiki and individual writing groups. "Engagement with other students" ($t_{50} = 2.16$, $p < 0.05$) and "cognitive engagement" ($t_{50} = 2.08$, $p < 0.05$) from student engagement ratings were the only two measures that showed a significant difference between groups. Attendance at tutorial sessions, but not grades, was marginally significant ($t_{50} = 1.88$, $p = 0.06$) according to the authors. Theoret and Luna found that writing through journals and writing through discussion boards are difficult to compare due to their fundamental differences, which they speculated has to do with the different audiences for the student writing. There were no differences in final course grades between the two groups (values not reported).

Phelps and Dostilio (2008) studied the student outcomes (project grade, final exam grade, and student reflection) to explore potential differences between a student-selected research project and one of two service-learning projects. GAISE recommendations were supported by either type of project and there were no statistically significant differences on the project or final exam scores. Student reflections, however, suggested statistical significance in their writing about "real world experience" (p-value = 0.019), "benefit to others" (p-value = 0.000), and "student development" (p-value = 0.005).

Zieffler and Garfield (2009) posed questions about student understanding of bivariational reasoning within the context of a course designed around the GAISE recommendations. Two sections of the same course covered the same materials, in the same way, with the same instructor but with two different sequences. No group

differences were detected but it was interesting that nearly all of the change in understanding took place within the first weeks of class when both sections covered sampling and exploratory data analysis but *not* bivariate data.

The review of research literature that references GAISE is almost exclusively piecemeal. The focus of journal articles is on one, or perhaps two, of the recommendations for teachers. Only the study of NSF-funded projects used the entire set of six recommendations and a single one specifically mentioned the goals for students. The proposed project intends to use both goals and recommendations as the conceptual framework for the construction of case studies.

Conceptual Framework

Figure 2. Guidelines for Assessment and Instruction in Statistics Education. This figure illustrates the distinction between content and pedagogy.

What it Means to be Statistically Educated				
Students should believe and understand why...	Students should recognize...	Students should understand the parts of the process through which statistics works to answer questions, namely...	Students should understand the basic ideas of statistical inference, including...	Finally, students should know...
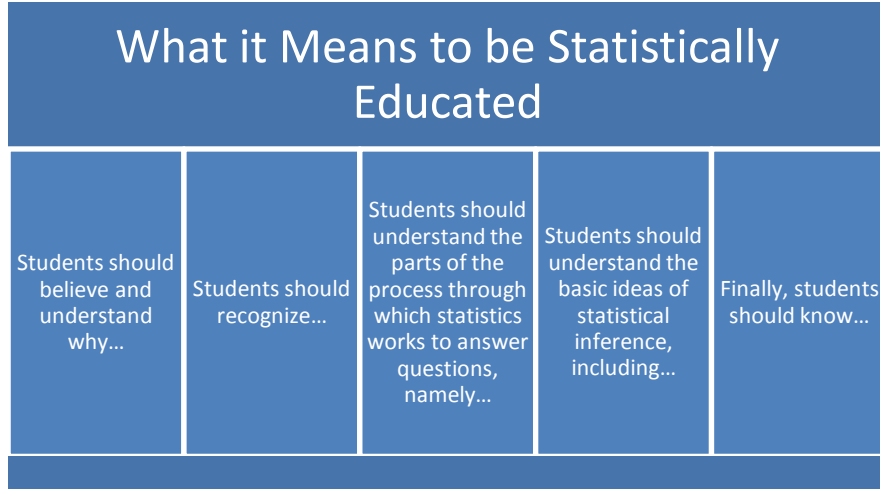
Figure 3. Five goals for students. This figure illustrates the categories for student outcomes for introductory courses.

Students should believe and understand why data is important, that variability is to be expected, how random sampling and random assignment are important aspects of study design, that association is not causation, and that statistical significance is not the same as practical importance, especially with large samples, nor that non-significance means there is no difference/relationship in the population, especially with small samples.

Students should recognize common sources of bias, the appropriate population to which results might generalize, when cause-and-effect conclusions are appropriate, and the difference between every day and statistical meanings for words like "normal," "random," and "correlation."

Students should understand the parts of the process through which statistics works to answer questions. This includes obtaining or generating data; graphing the data and knowing when that is sufficient; interpreting numerical summaries and graphical

displays; appropriate use of statistical inference; and communicating the results of a statistical study.

Students should understand the basic ideas of statistical inference. This includes the concepts of a sampling distribution and how it is important to making statistical inferences; statistical significance and *p*-values; and confidence intervals, specifically their interpretation of confidence level and margin of error.

Finally, students should know how to interpret statistical results in their context, how to critique news stories or journal articles, and when to get help from a statistician.

Recommendations for educators

How to become Statistically Educated

| Emphasize statistical literacy and develop statistical thinking. | Use real data. | Stress conceptual understanding, rather than mere knowledge of procedures. | Foster active learning in the classroom. | Use technology for developing concepts and analyzing data. | Use assessments to improve and evaluate student learning. |
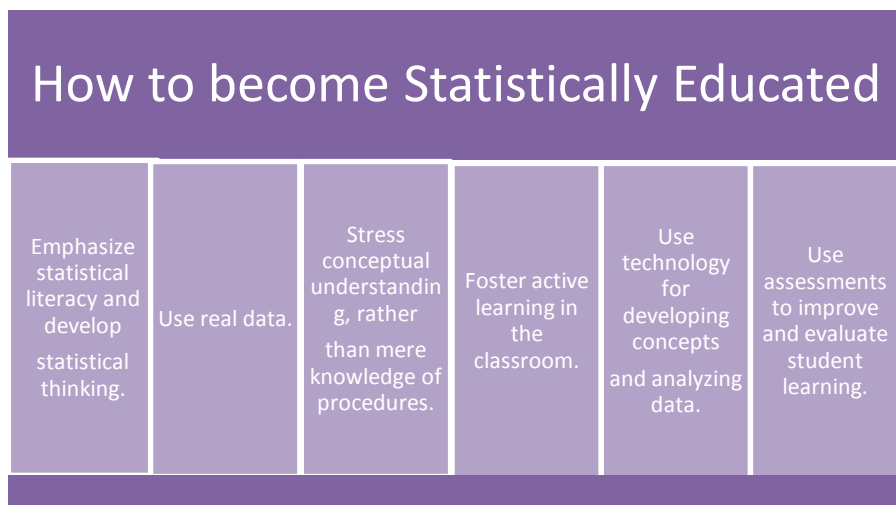
Figure 4. Six recommendations for teaching. This figure illustrates the suggestions for instruction to meet the student goals.

Statistics teachers are encouraged to model statistical thinking through well-articulated worked examples, and through use of technology for data management and analysis as well as inference and assumption checking; provide opportunities for practice including open-ended problems and projects where students much choose questions and techniques; and provide quality feedback through assessment.

Statistics teachers should search for and use data sets and summaries that are fresh and interesting to students; use class-generated data that is thoughtfully acquired to maximize in-class usefulness for illustration of many topics; require students to work with small sets of raw data but provide large sets electronically; and use data in multiple contexts (i.e. side-by-side boxplots and two-sample t tests).

Statistics teachers should consider the course goal not as coverage of particular methods but a set of underlying concepts. This change of perspective is likely to reduce the number of techniques introduced but allows for deeper understanding of key ideas. Similarly, use of technology for computation leaves more time to emphasize the interpretation of the result.

Statistics teachers should ground activities in the context of real problems; intermix lectures with activities and discussion; provide physical explorations and computer simulations; encourage prediction before analysis; allow students to suggest approaches to problems before procedures are introduced; provide formative feedback.

Statistics teachers should use technology not only for computation but also for visualization of concepts. Simulations provide opportunity to explore concepts. Technology also allows for multiple analysis techniques or graphical representations of data to explore conditions and presentations of data.

Finally, statistics teachers are encouraged to provide timely assessments with prompt feedback. The use of a variety of assessment options offers a more complete evaluation of learning. Interpretation and critique of news and graphs in the media assess statistical literacy, while open-ended tasks and projects assess statistical thinking.

About the Researcher

In the 23 years since completion of a bachelor's degree in mathematics (without a teaching certificate), I have worked for a non-profit health advocacy group, two hospitals, an automated recycling firm, and for construction/engineering consulting firms on exclusive contract to a consumer products manufacturer. The diversity of my employers is magnified by the multiple roles I filled at each. One such role—common to all but one position—is that of teacher. A master's degree in mathematics education finally gave me the credentials needed to make that *the* role I could fill.

Four years as an adjunct instructor of business statistics at Lakeland College included the opportunity to be contracted out to Bellin College of Nursing for an introductory statistics course. Returning to my master's institution, I was able to fill a sabbatical semester by teaching three sections of a Data Analysis course that was pre-requisite for application to the College of Education. Lack of opportunity for full time teaching and too many winters in Wisconsin encouraged me to seek employment in a warmer climate and pursue further education. The first step in that journey was an academic year of teaching statistics at James Madison University's College of Business and application to the doctoral program in mathematics education at the University of Virginia.

The choice of dissertation topic, like the literature reviewed above, is a convergence of two streams of interest: quantitative literacy in the workforce and statistics as an application of mathematics. My employment history outside of academia provides a real-world perspective on both. The coursework I have done during my

doctoral studies deepened my understanding of applied statistics while clarifying my

awareness of the importance of statistical literacy for *all* university students.

Chapter 3: Method

Research Design

In order to understand the similarities and differences of course objectives and implementation in undergraduate statistics courses across different academic departments, a thorough investigation of many subtle and inter-related factors is necessary. Yin (1994) suggests that case study research is well-suited to research interest in complex social and organizational phenomena. The delivery of introductory statistics courses by various academic departments is just such a complex phenomenon.

Stake (2000) suggests that multiple case studies build stronger understanding and more compelling evidence for findings by their discovery of patterns across the cases. Miles and Huberman (1994) describe the goal of analysis across multiple cases as explanation of how processes and outcomes are qualified by the different sets of conditions. This project asks questions that need to be answered by findings that compare patterns regardless of instruction (alignment with GAISE) and the influence of individual settings (academic departments).

Population and Sample

Central Virginia is well suited as the site for this study due to the concentration of institutions of higher education. The largest of these institutions is the University of Virginia, which serves as the principal research site. Within an hour's drive are four community colleges, two smaller public universities, several private liberal arts colleges,

and two for-profit institutions. Other than the for-profit institutions, each offers more than one course to which the Guidelines of Assessment and Instruction in Statistics Education should apply.

The University of Virginia offers nine undergraduate courses that have course titles or course descriptions identifying them as introductory, elementary, or first courses. These courses either have no prerequisites or require a particular mathematics course[3]; none require a previous statistics course. Two are offered as Applied Mathematics courses in the School of Engineering. In the College and Graduate School of Arts and Sciences (CGSAS), three courses are offered by the Department of Statistics, one by the Mathematics Department, and one each by the Psychology, Politics, and Sociology Departments.

The only exclusion criterion set for sample selection was a first time instructor. This eliminated a number of sections in the CGSAS where teaching assistants commonly teach introductory courses. Additionally, there were two refusals; one due to a perceived conflict of interest by the instructor with an administrative role and one by an adjunct faculty member with reservations about the time required to participate. There was also one non-responding instructor.

Expanding invitations to the surrounding colleges also met with several first time instructors and another non-responding instructor. Selection was further limited by time conflicts due to the travel requirements for data collection. A selective undergraduate institution was finally chosen for its combination of an experienced instructor, convenient time, and students of comparable backgrounds to those populating the courses being studied at the University of Virginia.

---

[3] Four high school credits in mathematics is a minimum entry requirement for admission to UVA.

A total of four courses became the case studies for this research project. Two are courses with calculus prerequisites; the other two had no mathematics pre-requisite beyond admissions requirements. The same two have instructors with terminal degrees in technical disciplines; the other instructors have terminal degrees in social science disciplines. One course has a small class size, one medium, one large, and one huge (well over 150 students). They all used course management systems and allowed the researcher access to the same documents that students could access. Graduate teaching assistants supported three of the four courses; the fourth had an advanced undergraduate student as a dedicated tutor. The cross-case analysis further details comparisons among the participating courses.

Data Sources

A key to strengthening the validity of the case study findings is the use of multiple sources of evidence (Miles & Huberman, 1994; Yin, 1994). Three main sources of data contribute to the case studies in the following chapter: printed documents (instructor-generated or formal publications), interviews with instructors, and classroom observations. These sources provide three perspectives on course characteristics, allowing for triangulation.

Syllabi, textbooks, and assessment documents were available for each case. Evidence of the course objectives and expectations for students come from the syllabi. Course content coverage is evident in syllabi, particularly in conjunction with the textbook and lists of reading/homework assignments. Quiz and exam documents inform the researcher of the importance the instructor attaches to specific areas of content.

The principal investigator attended the courses on at least seven occasions within a single semester to gain information on the enactment of the documented goals and expectations for the course. Observation of the first day of class was especially important as the foundation of the environment in which teaching and learning would take place throughout the semester. Two other preselected dates were observed based on topics shown to be persistently difficult for students according to the research literature (i.e., sampling distributions and introduction to inference). Additional observations took place to look for evidence of consistency in adherence to course objectives and expectations. The researcher maintained the role of an objective observer as far as possible in a social situation. Observations were video recorded, focusing on the instructor rather than the students. Transcriptions of the recordings supplemented the researcher's field notes and aided in data analysis. Brief, informal questions sometimes occurred in person or via email following a class observation and become part of the field notes.

Instructor interviews took place twice using a semi-structured protocol (see Appendix B), one before the semester began then at the end of the semester. Interviews were audio recorded and transcribed for coding. The purpose of these interviews was to gain insight into the instructors' experience with the course as well as their attitude and beliefs about the course's objectives and how the students are able/unable to meet them. The instructors' awareness of and compliance with the GAISE recommendations was investigated, implicitly at the beginning of the semester and explicitly at the end. The final interview also provided the instructor with an opportunity to reflect on the actual outcomes of the semester compared with his/her expectations at the outset.

Data Analysis

Marshall and Rossman (1999) point out that "data analysis is the process of bringing order, structure, and interpretation to the … data. It is messy, ambiguous. … It does not proceed in a linear fashion; it is not neat" (p. 150). They wrote about the analysis of qualitative data, but anyone who has ever dealt with raw quantitative data recognizes the sentiment as well. The key to useful analysis in case studies is careful organization before, during, and after data collection.

NVivo 9 is software specifically designed to organize qualitative data. Coding is significantly more efficient in this electronic environment. All sources of evidence are searchable and can be sorted by multiple criteria. The principal researcher purchased the software for use at home in addition to its availability at the University of Virginia's Scholar's Lab.

The six GAISE recommendations for teaching were the initial categories for coding:

1. Emphasize statistical literacy and develop statistical thinking;

2. Use real data;

3. Stress conceptual understanding rather than mere knowledge of procedures;

4. Foster active learning in the classroom;

5. Use technology for developing conceptual understanding and analyzing data;

6. Use assessments to improve and evaluate student learning.

Additional codes were needed to answer the first research question, "How do the introductory statistics courses offered by different academic departments define objectives and deliver instruction?" The GAISE list of goals for students were the starting place for codes, but other codes emerged to include some categorization of the ways in

which lecture content depends on the discipline with which the course is associated and the use of technology not related to conceptual understanding or data analysis.

Answering the complex, second question, "Are there commonalities that are sufficient for students in all classes to achieve the level of statistical literacy, reasoning, and thinking that the GAISE recommendations propose?" requires a pattern-matching approach. Cross-case analysis (Miles & Huberman, 1994; Yin, 1994) is intended to find patterns common to multiple courses (even if divergent from GAISE recommendations).

Trochim (1989) advocates for *pattern-matching,* where data is analyzed by comparing empirical patterns with predicted patterns (or alternatives). In this study, matching the course characteristics with the GAISE goals and recommendations is the primary analysis. When matches are not evident, alternatives are considered. This pattern-matching approach informs findings to both research questions.

 Participants had the opportunity for "member-checking" the final analysis. The advantage of this strategy is to test that researcher bias and data reduction have not interfered with 'truth' as seen by the participants (Krefting, 1999). Up to the date of this publication, participants have suggested only minor adjustments. Throughout the analysis phase of the study, a peer reviewer was consulted regarding coding and interpretation.

Chapter 4: Four Case Studies

The case analyses that follow are presented in two parts: description and pattern matching analysis. Each case is described in detail, including some of the known elements that make introductory statistics the "family of courses" (ASA, 2005, p. 10) that GAISE intends to address. The analysis of each case's match to the pattern set by GAISE is separated into two sections: Goals for Students and Recommendations for Teaching in the same way that the guidelines are divided.

In order to preserve the confidentiality of the participating instructors some details of interest are not as clear as a reader might wish. These details include the instructor's gender and department of appointment as well as the title of the course and the time of year for observations. Specifics about the instructor's experience and the type of students in the course are too important to the analysis to avoid mentioning but are intentionally vague.

The individual case analyses answers the first research question: "How do the introductory statistics courses offered by different academic departments define objectives and deliver instruction?" Chapter 5 will contain the cross-case analysis needed to answer the second research question: "Are there sufficient commonalities for students in *all* classes to achieve the level of statistical literacy and thinking recommended by the GAISE College Report?"

Case A – Statistics for Students in Technical Majors

The setting. There is a calculus prerequisite for the course. While students are not required to use calculus in determining probabilities, the textbook does demonstrate its use. The course is designed for students in science and technology majors. As with previous semesters, there are a small number of students majoring in non-technical areas that prefer this course to other options (Instructor interviews, pre- and post-semester). One such student, a business major, interrupted the pre-semester interview to obtain the instructor's signature on a form that would allow her enrollment.

There are just over 200 students enrolled in the three sections offered during the semester this study took place. Two of the sections are in the morning and the third in the early afternoon, all meeting on the same day and in the same classroom. The classroom has a capacity for 75 students and has rows of tables with a center aisle that approaches the projection screen. A whiteboard extends beyond the projection screen on both sides and occasionally holds announcements. Observations always took place in the afternoon class because it is the smallest section, ensuring that the observer had an unobstructed view of instruction. In every observation, the number of men exceeded the number of women by about a 2:1 ratio.

Professor A is an experienced instructor of calculus and probability/statistics. The professor has been at the institution for nearly a decade, first as adjunct faculty for the calculus sequence, then as an associate professor. While completing a doctorate in Industrial and Operations Engineering, Professor A had a teaching assignment at one of our nation's military academies followed by employment in the federal government and civilian companies prior to coming to this institution (Instructor CV). Teaching—and

now coordinating all instruction for—this course has been a main duty for both semesters

of the previous six academic years (Instructor interview, pre-semester).

Course design. During the summer prior to this study, Professor A participated

in a course design workshop offered by the institution's faculty development office

(Personal communication, pre-semester). The initial interest in attending the workshop

came from the professor's observation of student performance and "a vague sense that

there had to be a better way" (Personal communication, post-semester). According to the

related website:

> The design principles on which the [workshop] rests are grounded in the literature
> on course and syllabus design, educative assessment, active learning, and student
> motivation. Three components make our approach powerful: a taxonomy of
> significant learning, and the concepts of backward course design and integrated
> course design.

In the initial conversation regarding participation in this study, Professor A expressed

great enthusiasm for the redesigned course (Personal communication, pre-semester). The

course syllabus explains the new approach to the course structure:

> A minimal amount of time will be spent with lecturing. You will be provided a
> complete set of lecture notes in pdf form before class and will be expected to
> study these notes, augment/personalize them based on associated readings in the
> text book, and come to class ready to ask questions and discuss the contents of the
> notes. Many of the lessons will lend themselves to demonstrations/activities that
> will enhance your understanding of the material and add to your appreciation of
> what we are doing and the richness of its applications.

Until the course redesign, Professor A spent almost all of class time lecturing and waiting

for students to copy notes. "I used to give them only part of the slide ahead of time then

using the doc cam to reveal the rest during lecture. Some students refused to print it so

they were still feverishly copying the whole thing" (Instructor interview, pre-semester).

Implementation of preparedness quizzes using a classroom response system ("clickers") encourages students to take responsibility for their own learning as well as providing immediate feedback to keep them informed individually about their grasp of the basic concepts. It also gives the professor an aggregate view of the class's understanding before determining the class time needed on those concepts (Personal communication, pre-semester; Instructor interview, pre-semester; Syllabus).

> I envision going through some pages and ask a student to explain what my notes were trying to convey or just point out that a page was merely a summary of what was in some part of the book and, if there are no questions, just move on. Other pages will have some tough stuff and the notes will just be a place for a common ground for further discussion. (Instructor interview, pre-semester)

An online course management system is available to course participants, including the observer. All three sections had access to the same resources which included lecture notes, homework assignments with their solutions (after grading), project assignment with an example (after grading), post-quiz answers, and post-exam solutions. Students registered their clickers through the course management system and the professor also used it to send mass emails on a regular basis. The grade book was unevenly updated; exam and project grades were posted soon after due dates and points for clicker quizzes were entered weekly, but the traditional quiz and homework grades were withheld until late in the semester.

Lecture notes have titles that match the textbook chapters (see examples in Appendix C) and are followed almost linearly during class meetings. The course is segmented into three units that roughly align with the textbook chapters 1-4 (descriptive statistics, probability, and probability distributions), chapters 5-6 (confidence intervals and hypothesis testing), and chapters 7 & 9 (ANOVA and simple linear regression).

Some varieties of hypothesis testing were saved for the last week of class in order to address regression early enough to include in the course project and minimize the impact of waning attendance at the end of the semester (Instructor interview, pre-semester). The textbook material on multiple regression and quality control are not part of the course at this time, though they are both mentioned in the instructor's final interview as potential areas for expansion in future semesters.

Each section of the course is assigned one graduate assistant but Professor A pools their efforts as graders for exams and quizzes as well as "workshop" staff. The workshop functions as an open tutoring lab with specific hours when a statistics assistant is present (in addition to others assisting for various courses). Students typically request homework assistance at the workshop but sometimes need further explanations about concepts. The graduate assistants also proctor the two evening and the final exams. The particular graduate students assigned to the statistics course are offered the job based solely on their success in a single statistics class on their undergraduate transcript. Although not expected to attend lectures, they have access to the notes and homework solutions ahead of the students (Personal communication, post-semester).

Assessments. Four of the seven class observations begin with a "clicker quiz." This is the new tool for assessment of student engagement with the content in preparation for each class. "A two-question quiz (clickers) at the beginning of each class is just meant to test to see if they came prepared. No tricks or hard questions, just some basics that they should know from reading the book" (Instructor interview, pre-semester). Based on the interview, I expected to find these to be more about vocabulary or simple concepts; instead, those given early in the semester were mainly computations. For

example, the very first question presented to students for a clicker response asks for the mean of a distribution from the given probability mass function (Observation 2).

The instructor asked to see me briefly prior to the third class observation. During the preparations for the first exam, student complaints about the new course structure and Professor A's own disappointment with scores on the clicker quizzes led to an anonymous survey to clarify the issue(s). The students expressed their perception that the clicker quizzes were unfairly testing them on material not yet covered (Instructor interview, week 6). For the rest of the semester the clicker quizzes came after lectures.

More traditional paper-and-pencil quizzes are scheduled approximately every other week at the same time that homework sets were due. Students were given hard copy sheets with the quiz questions; solutions were posted online after each quiz. The second class observation was a quiz day. The class period had been quite full of new content and many students had difficulty completing the quiz before the end of the class time (Observation 2). At the final interview, Professor A mentioned that this was a lesson as one that did not go as well as expected: "There are so many things that are so fundamental that I decided that that would be better as two lessons this next go around."

Very few of questions across eight quizzes target conceptual understanding. Quizzes 1 & 6 contain exactly one conceptual question each:

> A list of 24 numbers has a mean of 52 and a median of 58. Suppose the ten smallest numbers are changed to smaller numbers and a new mean is calculated. What is the value of the mean after the change? (i) < 52 (ii) $= 52$ (iii) > 52 (iv) not enough info (Quiz1Soln)

> Give a <u>brief</u> interpretation of the interval you calculated in part (a), indicating the precise inference that your stated interval makes on μ. (Quiz6Soln)

Quiz 8 is almost entirely (8 out of 10 questions) conceptual, dealing with correlation and simple linear regression where calculations are time-consuming without a computer. For example, the final question asks students to complete the sentence "A confidence interval on the mean response will be smallest if the value of x is" by choosing "a) close to $\bar{x}$; b) far from $\bar{x}$; or, c) does not depend on $\bar{x}$."

Many of the other questions are entirely procedural, asking students to compute, calculate, or find specific values; a few are questions of definition or identification such as "Write a formula for the sample variance" (Quiz1Soln) or "If the P-value is 0.02, the null hypothesis is rejected at the 1% level, True or False" (Quiz7Soln). The quiz on hypothesis testing has questions that blend concepts and procedures: "…find the P-value" (procedural) "…then state your conclusion" (conceptual) (Quiz7Soln).

The exams are similarly heavy with computations. The final exam consists of 29 questions with varying point values. A little more than half (16) of them are strictly procedural, mainly computations, and are worth 55% of the grade. Six of the questions are strictly conceptual and contribute 13% of the grade. The remaining seven questions, worth 32% of the grade, either ask for an interpretation of completed calculations (as quoted above) or ask for a calculated value from partial information such as an incomplete ANOVA table.

Weekly homework assignments come from the even-numbered exercises in the textbook. The final answers are posted along with the assignment. The instructor explained that providing the answers prevented students from spending too much time following the wrong paths and not knowing they had erred early enough. Homework is

graded by an undergraduate who is given a solution key and told which intermediate steps must be present for the student to gain full credit.

The course project is motivated by the draft lottery for the Vietnam War, with some data and a link to more information provided. "This project, on a very small scale, will attempt to illuminate the process and subsequent analysis that was performed on the much larger and vastly more important lottery drawing during the Vietnam War" (ProjectDescription_v3). The actual work done by students, however, is couched in language regarding raffle drawings:

> We will compare the permutations produced from the four mixing/drawing strategies with other permutations that could have been drawn. Using the correlation coefficient as our test statistic, we will see if the results support our conjecture that the less mixed tickets will lead to a negative bias (indicating that the latter purchased tickets were generally drawn before the lower numbered tickets). Computed P-values will provide us with evidence about the significance of our results. (ProjectDescription_v3)

This is a partner project with self-selected pairs who do not have to be enrolled in the same section. The instructions suggest that working as a group is typical in technical fields. There are *very* specific guidelines about the format and page length. Professor A describes this type of rigidity as being common in requests for proposals or funding (Observation 5).

Case A – Statistics for Students in Technical Majors, Pattern Matching Analysis

Although Professor A was not aware of the *Guidelines for Assessment and Instruction for Statistics Education* (GAISE) (Instructor interview, post-semester), there is evidence to suggest that this course has some of the same goals for students and utilizes some of the recommended pedagogy. As mentioned before, the goals have much in common with modern textbooks and the recommendations for teaching are solidly

grounded in tenets of effective teaching. The lack of awareness of GAISE, therefore, does not preclude strong pattern-matching between the theoretical and the actual.

Goals for students. The observation protocol included a complete list of the goals as presented in the GAISE College Report (see Appendix A) and were simply checked off when the professor mentioned one during class time. Initial analysis of the coding on course documents through NVivo 9 revealed the same trend through frequency counts. This quantitative estimate demonstrates that Professor A's goals for the students are most aligned with items in the first, third, and fourth blocks of the GAISE list (see Tables 1, 3, & 4). The tables presented with each block give the frequencies of Verbal (observed lectures and interviews), Written (lecture notes, exam reviews, and syllabus), and Assessed (quizzes and exams) occurrences that match the individual goals. A more detailed analysis follows each table.

First block of goals. The goals in this block (see Table 1) relate to important concepts about what information statistical analysis can and cannot provide.

| Table 1 | | | |
First Block of Goals, Coding Frequencies			
Students should believe and understand why…	Verbal	Written	Assessed
Data beat anecdotes	0	0	0
Variability is natural, predictable, and quantifiable	4	15	0
Random sampling allows results of surveys and experiments to be extended to the population from which the sample was taken	2	3	0
Random assignment in comparative experiments allows cause-and-effect conclusions to be drawn	1	2	0
Association is not causation	1	1	0
Statistical significance does not necessarily imply practical importance, especially for studies with large sample sizes	0	1	0
Finding no statistically significant difference or relationship does not necessarily mean there is no difference or no relationship in the population, especially for studies with small sample sizes	0	1	0
Note: Cumulative frequency does not match NVivo count because some items matched multiple goals. Also, the Written and Assessed are undercounted because neither the textbook nor the final exam was available electronically.			

The goal for students to understand why data beat anecdotes is the only item in this block that received no coding. It is implied in the discussions of experimental design and inferential methods, but students are left open to the potential misconception that the absence of data for decision-making means that chance is the remaining influence, disregarding hearsay or misinformation.

Discussion of variability as natural, predictable, and quantifiable is found in a number of observations as well as the textbook reading and lecture notes that accompany them. Only the chapter/notes on classical probability neglects to mention variability explicitly in a statistical context. The first class of the semester, however, connects the

important concepts that "results will differ from one sample to [the] next … but variation can be modeled mathematically (based on probability concepts)" (Lecture1_Sampling&DescriptiveStatistics, p. 4; Observation 1).

For students who had already read the syllabus before the first class, they would have seen this connection expressed as the motivation for including the study of probability with statistics:

> We live in a world filled with uncertainty in which a lot of what happens to us (e.g., success/failure in school, careers, and decisions) is influenced by random factors. Probability is a means of capturing and analyzing events with uncertain outcomes and making wise choices in the face of uncertainty … we begin our course with basic concepts of probability not only for understanding in its own right but also for the foundation necessary to understand statistics. (Syllabus, p. 1)

The textbook includes a chapter on the "Propagation of Error" and is thus included in the class lectures. "Measurement is fundamental to scientific work. … Any measuring procedure contains error…,[which] is propagated from the measurements to the calculated value" (Navidi, 2011, p. 164). The use of repeated measures is useful to estimating the uncertainty within the measured values (Lecture3_Propagation of Error, p. 5). In addition to more subtle references to variability throughout the inference part of the course, this chapter emphasizes the goal that students should believe and understanding variation as natural, predictable, and quantifiable.

The other goals in this block receive less attention but are present in the course. On the first day of class, the lecture notes specify that statistical "methods involve design of experiments that allow reliable conclusions to be drawn from data produced" (p. 2) by sampling from a population of interest and that a "simple random sample is likely to representative (not biased) …, [which] guarantees statistically dependable results" (p.3).

Random assignment is mentioned in the lecture notes about analysis of variance as being "aimed at balancing nuisance variables" (Lecture9_ANOVA, p. 4), which had been described in the textbook section on correlation as "a third variable that is correlated with both of the variables of interest, resulting in a correlation between them" (Navidi, 2011, p. 514).

The definition of confounding (or nuisance) variables quoted above comes from a half-page subsection of the textbook titled "Correlation is Not Causation" (Navidi, 2011, p. 514). Lecture notes reiterate this as "correlation does not necessarily mean cause and effect" (Lecture7_CorrelationAndRegression, p. 7). The same lecture includes mention that a non-significant correlation does not mean that no relationship exists, only that a linear relationship is not supported by the data; there is no consideration of inadequate sample size nor is a similar statement made about failure to find statistically significant differences during hypothesis testing. The lecture notes on hypothesis testing does mention that "statistically significant results may not be 'important' results" (Lecture6_HypothesisTesting, p. 16) and uses a recent clinical drug trial as a real world example of an occasion when this was ignored.

Overall, Case A exhibits an uneven alignment of course material with the first block of GAISE goals for students. The alignment is strong regarding variability and randomization, with fewer opportunities to emphasize the difference between correlation and cause/effect. More discussion of non-significance would strengthen the course's alignment with the GAISE goals.

Second block of goals. The goals in this block (see Table 2) relate to appropriate interpretation of results from statistical analyses.

Table 2
Second Block of Goals, Coding Frequencies

Students should recognize…	Verbal	Written	Assessed
Common sources of bias in surveys and experiments	1	2	0
How to determine the population to which the results of statistical inference can be extended, if any, based on how the data were collected	1	2	0
How to determine when a cause-and-effect inference can be drawn from an association based on how the data were collected (e.g., the design of the study)	0	0	0
That words such as "normal," "random," and "correlation" have specific meanings in statistics that may differ from common usage	0	1	0

Note: Cumulative frequency does not match NVivo count because some items matched multiple goals. Also, the Written and Assessed are undercounted because neither the textbook nor the final exam was available electronically.

Data collection via a survey is not evident in the course. Two observations include discussions of sources of bias from experiments even though the accompanying lecture notes do not address the topic explicitly. In the first observation, Professor A asked students for suggestions about how sample selection may produce bias in the data collected and received several responses related to common sources of sampling or measurement errors (Observation 1). The final lecture on hypothesis testing covers t-tests on the difference of means with both dependent and independent samples. Again Professor A asks the students for suggestions about the advantages to paired data experiments and receives responses that display an understanding of confounding variables and the importance of study design on what inferences can be drawn (Observation 6).

The lectures on probability distributions include discussion of random number generators (CommonlyUsedDistributions, p. 71). This is the only place where the case

evidence mentions that the use of a specific word in everyday parlance could differ from its mathematical or statistical use. There was not an observation on the day of that lecture to know how much emphasis this instance received, but the textbook does not mention it at all and it is not repeated in any observation, other lecture notes or documents.

Case A offers only three instances of evidence to suggest that this course has goals aligned with the second block of GAISE goals. The evidence regarding study design is from informal discussion rather than explicit objectives for the course and the distinction of usage for the word "random" is not corroborated.

Third block of goals. The goals in this block (see Table 3) relate to procedures for obtaining and analyzing data with appropriate techniques and meaningful communication of the results.

Table 3 Third Block of Goals, Coding Frequencies			
Students should understand the parts of the process through which statistics works to answer questions…	Verbal	Written	Assessed
How to obtain or generate data	0	2	0
How to graph the data as a first step in analyzing data, and how to know when that's enough to answer the question of interest	1	6	1
How to interpret numerical summaries and graphical displays of data—both to answer questions and to check conditions (to use statistical procedures correctly)	3	3	1
How to make appropriate use of statistical inference	1	4	3
How to communicate the results of a statistical analysis	1	1	8
Note: Cumulative frequency does not match NVivo count because some items matched multiple goals. Also, the Written and Assessed are undercounted because neither the textbook nor the final exam was available electronically.			

The first chapter of the course textbook and, consequently, the first two lectures of the semester focus on obtaining data, graphing it, and creating numerical summaries from it (Navidi, 2011; Observation 1; Syllabus). These topics are revisited on several occasions throughout the semester, particularly in the context of interpretations needed for checking conditions before choosing appropriate testing procedures (Lecture6_HypothesisTesting). Creating a scatterplot from bivariate data prior to analysis of correlation or regression also depends on the ability to graph data and interpret it (Lecture7_CorrelationAndRegression). The final lecture in the first unit covered simulation and bootstrapping as means of generating data (Lecture4_CommonlyUsedDistributions; Navidi, 2011, p. 302-314).

Several assessments address the goals in this block directly. In Quiz 6 students are asked to "Give a brief interpretation of the interval you calculated in part (a), indicating the precise inference that your stated interval makes on μ." They are also asked to choose a precise interpretation of a p-value result in Quiz 7. The first exam exhibits the box plot of a skewed distribution and asks for a comparison between mean and median for the data.

Three of the seven course objectives listed in the course syllabus relate directly to this block of goals:

> 1. Interpret the meaning of data based on summary statistics and visual representations.
> 3. Describe a set of techniques for analyzing data, performing common statistical tests, estimating parameters, fitting data with functions, predicting values of variables based on models, and explaining variation.
> 7. Approach and solve real world … problems confidently using statistical techniques. This involves defining the problem, gathering information, identifying primary parameters, designing and conducting appropriate statistical experiments, analyzing data, evaluating findings, and presenting the solution in written form. (Syllabus, p. 2-3)

These three course objectives align with the GAISE goals for students to understand parts of the statistical process through interpretation, appropriate use of statistical inference, and communication of statistical analysis. The frequencies in Table 3 attest to the importance of these objectives.

   Fourth block of goals. The goals in this block (see Table 4) relate to important concepts needed for accurate interpretation of inferential analysis.

Table 4 Fourth Block of Goals, Coding Frequencies			
Students should understand the basic ideas of statistical inference…	Verbal	Written	Assessed
The concept of a sampling distribution and how it applies to making statistical inferences based on samples of data (including the idea of standard error)	3	7	1
The concept of statistical significance, including significance levels and p-values	1	13	0
The concept of confidence interval, including the interpretation of confidence level and margin of error	2	7	2
Note: Cumulative frequency does not match NVivo count because some items matched multiple goals. Also, the Written and Assessed are undercounted because neither the textbook nor the final exam was available electronically.			

   In addition to the course objectives quoted above that relate to inference, the syllabus also sets the expectation that students will "understand the logic of statistics. You will understand the conceptual and mathematical basis for the techniques of data analysis and representation, estimation, and hypothesis testing" (p. 2). Repeated insistence that the Central Limit Theorem is "the most important result in statistics" (Lecture4_CommonlyUsedDistributions, p. 52; Navidi, 2011, p. 290; Syllabus, p. 1) and appearance in multiple assessments (Quiz6Soln; Test2Soln) attests to the goal of having students understand how sampling distributions apply to making inferences.

More than half of the classroom lectures (23 out of 40) covered inferential topics (Syllabus, p. 4-5) with plenty of emphasis on the concepts of significance and confidence (Lecture5_ConfidenceIntervals; Lecture6_HypothesisTesting), suggesting the importance of these objectives. The lecture notes on hypothesis testing paraphrases the textbook's definition of a p-value as "the probability that a number drawn from the null distribution would disagree with H_0 at least as strongly as the observed value" of the statistic (Navidi, 2011, p. 398). Another perspective consistent between lecture notes and textbook is that α is the point at which H_0 is no longer plausible. Although more often couched in the vocabulary of plausibility and agreement/disagreement with H_0, the concept of significance levels and p-values is prominent in discussions of five different hypothesis tests of differences, regression, and factorial analysis (Lecture6_HypothesisTesting; Lecture7_CorrelationAndRegression; Lecture9_ANOVA).

The concept of the confidence interval is also covered by both the textbook and lecture notes (see example in Appendix C). During the last observation containing new material, Professor A demonstrated the construction of confidence intervals for the difference of two means when the variances are assumed to be equal and when that cannot be assumed. In each case, the final statements were "We are 95% confident that the true difference lies between here and here" and "We are 99% confident that the difference between the means is in this interval." The textbook is less precise in providing this final interpretation, generally stopping at a conclusion such as "the 99% confidence interval is … (520.12, 815.92)" (Navidi, 2011, p. 326).

In Case A the vocabulary of confidence intervals does not include "margin of error" in any of the collected documents or textbook. However, there is detailed

discussion of the meaning of an interval's width and the inverse relationship between

precision and confidence:

> • Narrower interval (e.g., $49 < \mu < 51$) – More precise inference – Less "confidence" that it captures the parameter
> • Wider interval (e.g., $47 < \mu < 53$) – Less precise inference – More "confidence" that it captures the parameter
>   (Lecture5_ConfidenceIntervals, p. 4)

Later in the lecture notes on confidence intervals, after discussion of determining a

sample size to achieve a specified width for the interval, the inverse nature of the

precision/confidence relationship is re-iterated with the conclusion that there "Must [be] a

trade-off or we can increase the sample size" (p. 16). The second exam assesses student

understanding of this relationship (Test2Soln, p. 2). Overall, there is strong alignment

between Case A and the fourth block of GAISE goals for students, allowing for

alternative vocabulary.

   Fifth block of goals. The goals in this block (see Table 5) relate to critical

thinking about statistical results.

Table 5 Fifth Block of Goals, Coding Frequencies			
Finally, students should know...	Verbal	Written	Assessed
How to interpret statistical results in context	1	3	2
How to critique news stories and journal articles that include statistical information, including identifying what's missing in the presentation and the flaws in the studies or methods used to generate the information	0	2	0
When to call for help from a statistician	0	0	0
Note: Cumulative frequency does not match NVivo count because some items matched multiple goals. Also, the Written and Assessed are undercounted because neither the textbook nor the final exam was available electronically.			

Professor A is more thorough than the textbook at providing a final interpretation of the statistical results framed in the context of the original questions. This discrepancy is most notable in the construction of confidence intervals as quoted in the previous section, where the textbook states the interval bounds but the professor references the parameter being estimated.

The syllabus specifies that one of the course goals is to "Be statistical critics and detectives: You will learn to question the characteristics, sources, biases, and implications of a set of data so you may intelligently evaluate statistical claims made in both professional literature (articles and conference proceedings) and the popular media (the press, advertising, books)" (p. 3). Professor A verbalizes this goal in the initial interview:

> I guess I would say I hope they have a better appreciation for probability and statistics. How it appears and is used in their daily lives – personal, professional – things that they read about … Can they tell the difference between good results and bad results? … I would hope that they gain an appreciation for some of the uses and misuses of statistics.

Table 5, however, suggests that there are no explicit efforts toward this goal. The two instances of coding captured in the table are the pre-semester statements quoted here. Lectures and the textbook do address the characteristics, sources, biases, and implications of a set of data (see the analyses on previous blocks) which give students the tools but not the practice for critique.

Professor A agrees that the students have the basics for meeting this objective but there "could be some interesting things to do, for discussion purposes in class; we could pose some situations and they would interpret that. That would be my next step, I think, to introduce some things like that" (Instructor interview, post-semester). The addition of such an activity or specific demonstrations of the impact of poor research design would

improve the implementation of this goal. Efforts to make critique a more obvious part of the course are likely to expose some situations where the help of a statistician is necessary, thereby introducing this goal to the course.

Recommendations for teaching. The observation protocol included a list of the six recommendations for teaching presented in the GAISE College Report (see Appendix A). At the end of an observation the session was rated—using a four-point scale from "not present" to "major part"—on the professor's inclusion of each recommendation. The Verbal column of Table 6 summarizes the frequency counts of observations where the recommendation was rated as having "part" or "major part" of the class. Initial analysis of the coding on course documents through NVivo 9 provide frequency counts on lecture notes, exam reviews, and syllabus in the Written column of Table 6. The counts in the Assessed column come from the project, weekly quizzes and the first two exams; the final and clicker quizzes were not available electronically for NVivo analysis.

Unlike the Goals for Students, the GAISE recommendations for teaching do not have a list of precise expectations to go with each section. There are instead examples of ways to include the recommendation for teaching, not all of which are useable in every course. For example, "Demonstrations based on data generated on the spot from the students" (ASA, 2005, p. 18) may be quite difficult in a large class and "Use a separate lab/discussion section for activities" (ASA, 2005, p. 19) may not be possible for small classes or at some institutions. The following analysis, therefore, demonstrates a wider range of evidence that matches or conflicts with each recommendation.

Table 6 Recommendations for Teaching, Coding Frequencies			
	Verbal	Written	Assessed
Emphasize statistical literacy and develop statistical thinking	5	28	14
Use real data	0	3	1
Stress conceptual understanding, not merely knowledge of procedures	4	10	11
Foster active learning in the classroom	4	2	0
Use technology for developing concepts and analyzing data	1	1	5
Use assessments to improve and evaluate student learning	2	0	0

Note: Frequency of Written occurances do not include the textbook.

Emphasize statistical literacy and develop statistical thinking. The evidence

from Case A fits well with the definition of statistical literacy used in the GAISE College

Report: "understanding the basic language of statistics… and understanding some

fundamental ideas of statistics" (ASA, 2005, p. 14). In the Goals for Students related to

understanding the process and the basic ideas of statistical inference, there is plenty of

evidence that Professor A expects students to be statistically literate at the end of the

course. "I would say so, based on the performance on the final exam … I think that's

reflected in the fact that a lot of students got through everything at 90% or above"

(Instructor interview, post-semester).

Developing statistical thinking, however, is lost in the effort to cover a breadth of

material. Among the suggestions for teachers to implement this recommendation is the

counsel to "Model statistical thinking for students, working examples and explaining the

questions and processes involved in solving statistical problems *from conception to*

conclusion" and "Give students plenty of practice with choosing appropriate questions

and techniques, *rather than telling them which technique to use* and merely having them implement it" (ASA, 2005, p. 15; emphasis added). Balanced by the strong emphasis on literacy, Case A demonstrates alignment with this recommendation.

 Use real data. The use of archival, classroom-generated, or simulated data in teaching provides authenticity and an opportunity to grapple with issues of data collection, and illustrates the connection to the problem context (ASA, 2005). The textbook for Case A includes many examples and exercises that refer to published articles (e.g., Navidi, 2011, p. 18, p. 326 & p. 419) but usually provides summary statistics rather than the data that produced them. The chapter on descriptive statistics does provide some data sets to demonstrate techniques (Navidi, 2011, p. 21 & p. 32) and a few small sets for exercises (Navidi, 2011, p. 45-47).

 Professor A points out, "In the project there's real data, presumably what is in the textbook is real data. I don't have real data so I use what works out" (Instructor interview, post-semester). The real data in the project is the simulations that the students generate to model sales of raffle tickets for analysis of four mixing/drawing strategies (ProjectDescription). The project is introduced with the controversy generated by the first draft lottery of the Vietnam War. Although actual data to replicate the randomization test of the draft lottery is not present (being much too large to even consider), it is a good example of statistical analysis being applied to real world problems. There are at least two other occasions when this type of "realness" is demonstrated: clinical drug trials to illustrate that statistical significance $\neq$ practical importance at the introduction to hypothesis testing and the example for testing the difference of two proportions (Lecture6_HypothesisTesting). However, references to

real-world use of data in decision-making may have more to do with fostering active learning by making the content relevant to the students' life experiences rather than being part of this recommendation.

As noted above, the textbook partially aligns with the GAISE recommendation to use real data, but most of the lecture material and assessments do not. The notable exception is when assignments include student-generated data through simulations such as the project:

> To simulate the scenario of ten people each purchasing 4 raffle tickets, the tickets should be dropped into the hat in groups: ticket numbers 1- 4 followed by numbers 5-8, …, and finally ticket numbers 37-40. This same sequence will be performed four times followed by the Mixing/Drawing Strategies described below. In each of these experiments, all tickets need to be drawn and the resulting sequence of ticket numbers needs to be recorded. (ProjectDescription, p. 4)

In the case of quizzes and exams, there is a time constraint that may prevent the use of real data sets. Homework assignments, however, might include some of the exercises referencing real data rather than contrived data. For example, the fourth homework assignment includes an exercise that asks, "Estimate the [parameter] and find the uncertainty in the estimate" (Navidi, 2011, p. 178) when the same page of the textbook has an unassigned exercise that refers to published results but asks the same question. Incorporating more examples that explicitly reference published statistical analysis in the lecture—even without discussion of that fact—is another small adjustment to the course that would improve its alignment with GAISE.

Stress conceptual understanding rather than mere knowledge of procedures. Case A demonstrates the blending of conceptual and procedural considerations during class time. The inferential topics are introduced with a clear pattern of "this is *why* we

want to do this," "this is *how* we do this," and "this is what our calculations *mean*" (Lecture5_ConfidenceIntervals; Lecture6_HypothesisTesting; Lecture7_CorrelationAndRegression). Like the earlier topics, however, they are assessed by quizzes and tests with heavier weight on the procedural, though some conceptual questions are included. On the other hand, homework assignments for the early topics focus more heavily on the conceptual but favor the procedural after the introduction of inference. The blending is inconsistent between classroom discussion and subsequent assessments but still demonstrates the importance of both in Case A.

NVivo coding shows some overlap of this recommendation and "emphasize statistical literacy." Most particularly this happens at the introduction of confidence intervals and hypothesis testing. This overlap is illustrated by the discussion of width in regards to confidence intervals (quoted on page 56) as well as a similar summary slide for hypothesis tests:

- Need to decide what level of disagreement, measured by the P- value, is great enough to render H_0 implausible
- The smaller the P- value, the more certain we can be that H_0 is false
- The larger the value, the more plausible H_0 becomes BUT we can never be certain H_0 is true
  (Lecture6_HypothesisTesting, p. 14)

Separating the objective "understanding some fundamental ideas of statistics" (ASA, 2005, p. 14) from the objective "stress conceptual understanding" (ASA, 2005, p. 17) is difficult to do when fundamental ideas are first presented.

The GAISE College Report links this recommendation with the idea of knowing important concepts well so that learning additional procedures are readily accomplished in a second course. Professor A recognizes this connection, retaining the procedures in

the course but putting them in the last week of the semester with the following

explanation:

> Today we'll go over two special cases of confidence intervals which becomes a nice, natural review of the chapter 5 material. On Wednesday and Friday we'll cover four miscellaneous cases for hypothesis testing. These topics are in the "cookbook" realm where you already know the basics and we're just going to see some different formulations. … It will be a refresher for the procedures of hypothesis testing. (Observation 6)

Overall, there is moderate alignment with this GAISE recommendation that could

be strengthened by further condensation of the "cookbook" procedures that fill the final

week of the semester. Spending part of that week on how to choose a technique would

also develop statistical thinking, simultaneously strengthening alignment with GAISE's

first recommendation for teaching.

Foster active learning in the classroom. The introduction of a classroom

response system facilitated an increase in active learning this semester. More time

devoted to working exercises in class this semester brought this course into alignment

with more than one of the GAISE recommendations.

> The think-pair-share strategy worked very well. It went especially well when done with the clicker response. … Basically, in the past, I would try to teach through my notes to be sure the students had the coverage. Now I am sort of giving them that, then having the opportunity to augment that in class. It freed up the time from taking notes, copying down the stuff I had; it allowed me to introduce these additional practice problems which I had not had time to do before. (Instructor interview, post-semester).

Professor A also made an effort to draw students into classroom dialogue. One

expectation for the redesigned course was that students would be prepared to answer

questions from their reading when called upon in class (Instructor interview, pre-

semester). This expectation met with mixed results, as seen in the second observation.

Students were using tent cards to display their names to the professor and were called by

name at three points in the lecture. Invariably the first student called on had no response or only a partially correct response. A second student either volunteered an answer or gave a better response after a hint from the professor. On two of these attempts, a third student had to complete his/her peers' thoughts to satisfy the professor. Later in the semester, Professor A does not call on particular students but poses questions to the full class and gets volunteers; in some cases, they are reluctant volunteers (Observation 6).

Use technology for developing conceptual understanding and analyzing data.
Professor A uses technology in three distinct ways, the first of which also addresses the recommendations for active learning in the classroom and using assessment to improve and evaluate student learning. As mentioned above, the use of a classroom response system, or clickers, is a new element in this course and Professor A plans to continue using them for preparedness quizzes and think-pair-share activities to further conceptual understanding (Instructor Interview, pre- & post-semester).

The second use of technology also focuses on conceptual understanding. Throughout the semester, lecture notes reproduce or supplement the textbook's use of graphical displays of data to emphasize or illustrate concepts (e.g., Lecture1_Sampling and Descriptive Statistics, Lecture4_CommonlyUsedDistributions, & Lecture7_CorrelationAndRegression). These are, of course, static representations (see example in Appendix C). Professor A may find that the reduction in lecture and note-taking time that this semester's adjustments have provided may also allow for the next step in technology inclusion by providing real-time, dynamic demonstrations on some occasions as aids to visualization of concepts as GAISE suggests.

A third use of technology is for the analysis of data through the use of statistical software. The textbook includes output from analyses run in the statistical software package Minitab to prepare students for work with larger data sets that require software assistance (e.g., Navidi, 2011, p. 401 & 553). Several homework problems (e.g., HW06_Assignment & HW09_Assignment) and the course project required student to use Minitab for analysis. Professor A says, "I provide them with a user guide and encourage them to use it for the descriptive[s] … I lead them by the hand in the simulation exercises: step 1 – do this; step 2 – do this. They have lots of [written] guidance" (Instructor Interview, pre-semester). As with the graphical displays, to strengthen the course's alignment with this teaching recommendation from GAISE live demonstrations may find a place in class time for future semesters.

Use assessments to improve and evaluate student learning. Individual feedback is not available for most assessments in this course. "With the number of students, it would just be too unwieldy to put notes on them" (Instructor interview, pre-semester). Complete solutions to exam questions are available in the course management system, in addition to a quick review in class where questions are welcome (Observation 3). Homework and quiz solutions are also available online for students who take the initiative to review them.

The use of clickers in the classroom supports Professor A's efforts to use assessments to improve and evaluate student learning. This is mainly due to their two-way value as an evaluative tool for both students and instructors. The feedback on the preparation quizzes and the think-pair-share activities is both "useful and timely," which the GAISE College report deems "essential for assessments to lead to learning" (ASA,

2005, p. 21). As a new tool in the course, the Professor found room for improvement by the end of the semester. "I think I need to take one more minute and show the students the histograms. They individually know whether they got it right or wrong, but to see the histogram…gives them some peer assessment that I did not share with them all the time. That would give them useful feedback" (Instructor interview, post-semester).

Summary. The pattern-matching analysis of Case A demonstrates a strong match with some of the GAISE Goals for Students and mention of nearly all of the goals. Emphasis on variability as being natural, predictable, and quantifiable provided the largest body of evidence for a single goal. The third and fourth blocks of goals—relating to analytical procedures and conceptual understanding needed for careful interpretation of statistical analysis, respectively—showed the most thorough matching between Case A's goals and those of GAISE with evidence for each of the eight individual goals. Only three of the GAISE goals had no matches in Case A: Data beat anecdotes, How to determine when a cause-and-effect inference can be drawn, and When to call for help from a statistician.

Of the GAISE Recommendations for Teaching, Case A shows convincing evidence of emphasizing statistical literacy and developing statistical thinking as well as stressing conceptual understanding over mere knowledge of procedures. The introduction of a classroom response system ("clickers") in the observed semester produced evidence that Professor A strives to foster active learning in the classroom and use assessment for learning. Using real data and using technology for developing concepts and analyzing data are GAISE recommendations with little evidence in Case A, but they are not entirely neglected.

Case B – Statistics for Students in Business Majors

The setting. This course is designed for students in business majors. There is a calculus prerequisite but no use of calculus in lectures, course documents, or in the textbook. "I don't have the expectation that they can use the techniques of calculus but they should have a familiarity with math. They should have a certain level of confidence with math, thinking mathematically and doing mathematical problems" (Instructor interview, pre-semester). Although not an official institutional prerequisite, the course syllabus specifies that "some in-class examples will be presented using Microsoft Excel…You will be required to understand aspects of this software that are discussed in class, such as the interpretation of output…Previous experience with Excel is recommended" (p. 2-3). Professor B acknowledges that "some people use calculators but I want something that everybody can use. I asked them for Excel familiarity" (Instructor interview, pre-semester).

There are 453 students enrolled in the three sections offered in the observed semester. One section is a mid-morning class, the other two meet in the early- and mid-afternoon. The afternoon classes meet in the same auditorium-style classroom but on different days. The morning class is held in a slightly smaller (245 v. 300) auditorium in a different building. Students can attend any of the three sessions regardless of which they registered. Observations took place in the early-afternoon class because it was the third time that Professor B presented the material as it was in Case A. It was also the session recorded and available to students for the rest of the semester through Blackboard Collaborate. There is not a visually striking difference in the numbers of men and women enrolled in the course.

This is the second time that Professor B has taught this course at this institution. Three faculty members rotate the duty to teach this large introductory course. "There are two of us now who have taught it before. I shouldn't say that – there are two of us who have taught it before *and* are interested in teaching it again" (Instructor interview, pre-semester).

Course design. An online course management system is available to course participants, including the observer. All three sections have access to the same resources which include the professor's lecture notes with accompanying presentation slides and Excel examples. Review notes and examples for each exam became available as exam dates approached. The questions and an answer key are accessible after each exam. Homework assignments come from an online database which the professor set to individualize for the students by using random values within a range, and most questions allow for multiple attempts. Recordings of the virtual lecture are available in the course management system. The discussion board is set up to provide administrative and technical support as well as general questions about content. Raw scores for exams are in the online grade book a few days after each exam, then adjusted as the final exam nears.

Lectures follow the sections of the textbook linearly, though the lecture notes are organized by "topic" rather than chapter. Twelve topics align with the first eleven chapters of the textbook (Topic_Schedule; Moore, McCabe, Duckworth & Alwan, 2011). Lecture notes have subsections labeled in almost perfect alignment (skipping an optional section in chapter 7) to the chapter sections; the lecture slides refer to the chapter sections by name and number (see example in Appendix C). The first exam assesses student learning about data, data collection, study design, random variables, and probability. The

second exam assesses the basics of inference: confidence intervals and significance tests. The final exam includes all topics, with simple linear and multiple regressions as the only new content assessed.

Case B has eight graduate students assigned to act as recitation leaders, homework supervisors, and as help lab staff. The help labs are open sessions meant to give students an opportunity to ask questions regarding homework problems and study assistance. Face-to-face and virtual help labs are entirely the responsibility of the graduate students. An informal conversation with a student suggested disparity in the quality of the help received during labs: "It depends on which TA you get. Some just tell you what formula to use but some will talk about why you use that one, which is more helpful" (Personal communication, week 7).

At the beginning of the semester, Professor B lectured for the entire class session on the first class of the week; one of the two recitation leaders would use the second for work on exercises, encouraging students to work together in small groups. In the fifth week of the semester, Professor B took over responsibility for the recitations, using the first part of the class to complete or review lecture material and working with the graduate student to implement the exercises. The change was precipitated by a missed session when a recitation leader overslept, in addition to "some student complaints about lack of access" to the instructor (Personal communication, week 7).

Assessments. Twelve sets of homework, totaling 280 exercises, and 102 exam problems are the only assessments in this course. All of the exam questions are in the forced-choice format as are some of the homework exercises. The overall distribution of

exam questions is 50% procedural, 33% conceptual, and 17% require a conceptual understanding in order to complete a calculation.

Some of the homework exercises are sets of four statements for students to label as true or false and there are a couple of matching exercises. The rest of the exercises have text fields for numerical responses. The homework questions are more procedural than conceptual. The earliest sets include a few vocabulary-based exercises as either forced-choice or true/false. Some questions require conceptual understanding in order to complete multi-step calculations. An example of the latter follows:

> Suppose that the mean score of an exam was 75 when 34 students took it on time with a standard deviation of 1.6. A makeup of the same exam is given to 5 students. The retakes averaged a score of 82 with a standard deviation of 3.1. What is the average of the test scores?

Case B – Statistics for Students in Business Majors, Pattern Matching Analysis

Professor B had not heard of GAISE before being asked about it in the post-semester interview. The pattern matching analysis that follows, however, reveals some important commonalities in the objectives and pedagogy for the course and what GAISE recommends for all introductory statistics courses.

Goals for students. All of the goals listed in the GAISE College Report were included in the observation protocol (See Appendix A) for tracking the verbal evidence of Case B's inclusion of those goals. Together with the frequencies of NVivo 9 coding on interviews, they make up the counts in the Verbal column of the tables that accompany the analysis. Counts in the Assessed column come from the coding of homework and exams. The Written column frequencies come from NVivo coding on lecture notes, the syllabus, and recitation activities. The textbook was not coded but does provide some of the supporting evidence in the analysis report.

Tables 7 through 11 provide initial indications of where Case B's objectives are most in line with those of GAISE. Blocks three (Table 9) & four (Table 10) received the most attention during coding and the detailed analysis confirms this preliminary assessment of the areas with the most evidence of alignment.

First block of goals. Important ideas about what information statistical analysis can and cannot provide are included in this first block of goals (see Table 7).

Table7 First Block of Goals, Coding Frequencies			
Students should believe and understand why…	Verbal	Written	Assessed
Data beat anecdotes	3	1	0
Variability is natural, predictable, and quantifiable	7	5	1
Random sampling allows results of surveys and experiments to be extended to the population from which the sample was taken	0	3	1
Random assignment in comparative experiments allows cause-and-effect conclusions to be drawn	0	2	1
Association is not causation	2	1	1
Statistical significance does not necessarily imply practical importance, especially for studies with large sample sizes	2	1	1
Finding no statistically significant difference or relationship does not necessarily mean there is no difference or no relationship in the population, especially for studies with small sample sizes	1	0	1
Note: Cumulative frequency does not match NVivo count because some items matched multiple goals. Frequency of Written occurances do not include the textbook.			

The first meeting of the semester introduced the textbook's definition of Statistics: "the science of data" (Moore, et al., 2009, p. 4; Lecture Intro, slide 4; Observation 1). Professor B spends 11 minutes of the 30 minutes available for non-

administrative topics emphasizing the importance of data, providing examples from the internet across many disciplines (Observation 1). Some of this discussion addresses the issue of "context" for the data, "How are the data…Produced? Collected? Organized? (Lecture Intro, slide 13).

The natural occurrence of variability in collected data also receives attention in the first lecture (Lecture Intro; Observation 1). During the description about the importance of data, the professor offers multiple definitions of statistics that emphasized data but have also mentioned variability often enough to be part of the suggestions students provide for the in-class composition of a more complete definition of statistics (Observation 1). The lecture notes anticipated this, including a slide with the title "Variation Breeds Uncertainty" that warns the audience "Variation is everywhere" as well as assuring them that "Statistics provides tools for dealing with variation and uncertainty" (Lecture Intro, slide 14).

The topic of variability is seen again during lectures on sampling distributions and probability, particularly in relation to random sampling (Lecture Topic 3). An Excel demonstration by the professor provides a real-time opportunity to see how a sample statistic can vary (Excel Demo 3; Observation 2). In the homework set on sampling distributions, one true/false question assesses student understanding of the relationship between variability and sample size: "As sample size increases, statistics become less variable" (Homework03). The final exam includes a similar question.

Topic 2 in the lecture notes and chapter 3 in the textbook cover both random sampling and random assignment. The lecture notes contrast probability sampling—simple random and stratified—to biased sampling techniques, after reminder definitions

of population and sample that connect them through inference. There are several "What is the population?" exercises in the textbook (e.g., Moore, et al., 2009, p. 191) but a single question on the first exam is the only formal assessment of student understanding that takes place in Case B. Random assignment is part of the lecture on experimental design as a way to eliminate biased results (Lecture Topic 2). The textbook chapter describes the difference between observational studies and experiments, concluding, "When our goal is to understand cause and effect, experiments are the only source of fully convincing data" (Moore, et al., 2009, p. 177). Homework02 rephrases this in a true/false statement.

The true/false section of Homework02 also includes the statement "Association does not imply causation." A similar statement appears in the lecture notes for Topic 2 with additional information about lurking variables. The textbook is very clear about both the temptation and the inappropriateness of assuming that a correlation is evidence for cause and effect:

> When we study the relationship between two variables, we often hope to show that changes in the explanatory variable *cause* changes in the response variable. But a strong association between two variables is not enough to draw conclusions about cause and effect. Sometimes an observed association really does reflect cause and effect. [omitted example] In other cases, an association is explained by lurking variables, and the conclusion that *x* causes *y* is either wrong or not proved. (Moore, et al., p. 143)

An activity in the recitation session asks students to critique a causal claim that depends on evidence of association (Activity Topic 2; Recitation 1).

Lecture notes for Topic 7 draw the distinction between statistical significance and practical importance and this is reiterated in the first observed recitation session (see example in Appendix C). The textbook—but not the lecture notes—specifies that "When

large samples are available, even tiny deviations from the null hypothesis will be significant" (Moore, et al., 2009, p. 399). Understanding of this facet of interpretation of statistical analysis is assessed with a sequence of true/false statements in Homework07:

1. Lack of significance implies that H_0 is true.
2. A statistically significant result is always practically significant.
3. Due to the common usage of $\alpha = .05$, there is a large practical distinction between the P-values 0.049 and 0.051.
4. A good way to help determine if an effect is practically significant is to plot the data.

The first statement in the above list addresses the last goal in this block about the appropriate interpretation of non-significant results of a hypothesis test. This is also addressed on the same slide in Topic 7, which distinguishes between significance and importance. As with the impact of large samples on significance, the textbook alone mentions that small samples may be "insufficient to detect the alternative" (Moore, et al., 2009, p. 399).

Every goal in this first block is evident in Case B, demonstrating alignment with GAISE.

Second block of goals. The appropriate interpretation of results from statistical analyses is the theme of the goals in this block (see Table 8).

Table 8 Second Block of Goals, Coding Frequencies			
Students should recognize…	Verbal	Written	Assessed
Common sources of bias in surveys and experiments	1	3	2
How to determine the population to which the results of statistical inference can be extended, if any, based on how the data were collected	0	0	1
How to determine when a cause-and-effect inference can be drawn from an association based on how the data were collected (e.g., the design of the study)	0	1	0
That words such as "normal," "random," and "correlation" have specific meanings in statistics that may differ from common usage	1	2	1
Note: Cumulative frequency does not match NVivo count because some items matched multiple goals. Frequency of Written occurances do not include the textbook.			

Chapter 3 of the textbook addresses the production of data and the sections on designing samples and experiments covers sources of bias (Moore, et al., 2009). The Topic 2 lecture follows suit:

Other sources of bias
- **Under-coverage** in the population list.
- **Non-response** of sampled individuals.
- Inaccurate responses of the respondent (**response bias**).
    o May be unintentionally encouraged by the interviewer.
- Poor **questionnaire design** and wording. (slide 27)

Student understanding of these potential sources of bias is assessed by a set of true/false statements in the homework set (Homework02) and a single identification question on the first exam:

A sampling study intends to generalize results to all residents of a certain town, but a simple random sample is collected only from those residents who are registered to vote. The bias in this setup is due to:
    a. Probability sampling using unknown selection probabilities.
    b. Non-response of the sampled individuals.
    c. Under-coverage of the population list.
    d. Voluntary sampling (MT1_Exam)

Determining the population to which inference is appropriate based on how the data were collected is assessed in the first exam (MT1_Exam), although there is no verbal or written evidence outside the textbook, as already mentioned above. Also previously mentioned is the recitation activity discussing the fallacy of inferring a cause-and-effect relationship based on a correlation (Activity Topic 2).

Chapter 4 of the textbook covers probability and probability distributions (Moore, et al., 2009). The lecture that aligns with this assigned reading emphasizes the mathematical meaning of the word "random" and its relationship to probability.

> Randomness and probability
> Observations of **random phenomena**:
> - Patterns emerge "in the long-run" after many repetitions of a chance-happening.
> - Short-term patterns are unpredictable.
>
> **Probability** attempts to describe the long-term patterns of random phenomena (Lecture Topic 3, slide 11)
>
> Often when we think of chance happenings or random phenomena, we think of things that we might describe as unpredictable, chaotic, structureless, patternless … something like that. Something with no form to it. But an interesting thing to observe with chance happenings is that when you observe it over and over again repeatedly, like I did with that sample earlier, you'll start to see that patterns do emerge and there is a certain structure. (Observation 2).

There is no evidence that other words with nuanced meanings that differ from the everyday meaning receive similar attention.

In this block every goal appeared at least once among the case evidence. The goal of recognizing common sources of bias has all three types of evidence: verbal, written, and assessed.

Third block of goals. The goals in this block (see Table 9) relate to procedures for obtaining and analyzing data with appropriate techniques and meaningful communication of the results.

Table 9 Third Block of Goals, Coding Frequencies			
Students should understand the parts of the process through which statistics works to answer questions…	Verbal	Written	Assessed
How to obtain or generate data	2	2	1
How to graph the data as a first step in analyzing data, and how to know when that's enough to answer the question of interest	2	3	0
How to interpret numerical summaries and graphical displays of data—both to answer questions and to check conditions (to use statistical procedures correctly)	1	2	8
How to make appropriate use of statistical inference	5	7	9
How to communicate the results of a statistical analysis	4	0	2
Note: Cumulative frequency does not match NVivo count because some items matched multiple goals. Frequency of Written occurances do not include the textbook.			

The first class of the semester includes a lengthy discussion of data mentioned as part of the evidence for the first block of goals. In that lecture ProfessorProf. B shares web links to national and international data as well as a collection of sports data (Lecture Intro, slide 5; Observation 1). Topic 2 describes the difference between observational and experimental studies, emphasizing the usefulness of the latter in making inferences. Chapter 3 of the textbook is appropriately titled "Producing Data" (Moore, et al., 2009). Both midterm exams include a question about the advantage of data from an experiment (MT1_Exam; MT2_Exam).

Graphing data as a first step of analysis is mentioned in connection with correlation and regression (Lecture Topic 2; Lecture Topic 12; Observation 6).

"Scatterplots provide a good 'first look' at the data" (Lecture Topic 12, Observation 6).

Interpreting numerical summaries (e.g., correlation coefficient) and graphical displays

(e.g., histograms) are topics early in the course (Lecture Topic 2; Lecture Topic 3). The

first exam includes four questions regarding interpreting numerical summaries

(MT1_Exam). The one that comes closest to discussion of a graphical display provides a

mean and median for a data set and asks the students "A mean-median comparison tells

us that the data are: (a) Multi-modal (b) Right-skewed (c) Left-skewed (d) Symmetric"

(MT1_Exam, question 4). Homework 02 also has eight true/false statements and a two-

part question on the use of a prediction equation to assess student use of numerical

summaries.

Issues of conditions/assumptions and robustness to violations are included in

lectures at the introduction of new inferential procedures. Each procedure has a "recipe

slide" such as this one:

One-sample t test

□ **Assumptions:** SRS of size n from a Normal population

□ **Hypotheses:** H_0: $\mu = \mu_0$ *versus* a one- or two-sided H_a

□ **Test statistic:** $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$

□ **P-value:**
  $P(T \geq \text{-}t)$ for H_a: $\mu < \mu_0$
  $P(T \geq t)$ for H_a: $\mu > \mu_0$
  $2P(T \geq |t|)$ for H_a: $\mu \neq \mu_0$
  where T is $t(n-1)$

Figure 5. Example of a "recipe" slide.

The textbook version of the "recipe" is a shaded box that uses text to describe the

assumptions and hypotheses and adds graphical depiction of the rejection region(s) (e.g.,

Moore, et al., 2009, p. 428). Activity Topic 2 begins with scenarios for which students

must apply what they know about study design and assumptions for procedures to choose the appropriate test (Recitation 1). Homework 9 asks students to "match each experiment below with the correct formula for its analysis," assessing a wider variety of procedures than the recitation activity.

The evidence regarding the communication of results overlaps with the interpretation of significance and confidence in the fourth block, therefore, this goal is analyzed below.

Case B contained widespread evidence for goals related to student understanding of statistical processes. Appropriate use of statistical inference and interpretation of numerical summaries were particularly evident.

Fourth block of goals. Important concepts needed for accurate interpretation of inferential analysis are the goals in this block (see Table 10).

Table 10 Fourth Block of Goals, Coding Frequencies			
Students should understand the basic ideas of statistical inference…	Verbal	Written	Assessed
The concept of a sampling distribution and how it applies to making statistical inferences based on samples of data (including the idea of standard error)	4	12	3
The concept of statistical significance, including significance levels and p-values	4	16	9
The concept of confidence interval, including the interpretation of confidence level and margin of error	2	12	9
Note: Cumulative frequency does not match NVivo count because some items matched multiple goals. Frequency of Written occurances do not include the textbook.			

The concept of sampling distributions gets a brief introduction in Chapter 3 of the textbook and Topic 3 in the lectures. The textbook's definition—"The sampling distribution of a statistic is the distribution of values taken by the statistic in all possible samples of the same size from the same population" (Moore, et al., 2009, p. 213)—is

79

followed by descriptions of shape, center, and spread without specific reference to "standard error" or the Central Limit Theorem.

Chapter 4 in the textbook and the Topic 4 lectures (on random variables and probability) culminate in more formal statements about the characteristics of sampling distributions with reference to the Central Limit Theorem. An Excel spreadsheet is used to demonstrate the "samples of the same size from the same population" (Moore, et al., 2009, p. 213) using repeated sampling and graphical displays (Topic 03 Examples). Appendix D shows a later example that was built on the repeated samples begun at this point of the semester. This spreadsheet is available to the students via the course management system and they are encouraged to observe repeated sampling on their own.

While there are a few homework questions (e.g., Homework 03) to assess student understanding of sampling distributions, there is only one exam question:

> When planning a sampling study, an effective way to reduce variability in the sampling distribution of a statistic is to...
> a. randomize the allocation of subjects to treatments.
> b. eliminate over-coverage of the population.
> c. increase the sample size.
> d. eliminate lurking variables.
(FE_Exam, question 24)

Confidence intervals and significance tests are both introduced using the case of a single mean in Chapter 6 and its accompanying lecture (Moore, et al., 2009; Lecture Topic 6). The components of these two types of inference named in this block of goals are all defined during the initial exposure. Repeated use and interpretation of "significance level," "p-value," "confidence level," and "margin of error" (in addition to "standard error" regarding the sampling distribution being referenced) provide many opportunities for students to grasp these concepts. Homework and exam questions are

heavily focused on calculating values (16 of the 33 questions on the second exam, which

assesses inferential ideas) but the second exam includes two questions specifically to

assess the conceptual understanding:

> Which statement below reflects a correct interpretation of a confidence interval?
> a. The formula used to calculate the upper and lower bounds of a 95%
> confidence interval for μ would, in the long run, yield an interval that
> includes μ 95% of the time.
> b. The formula used to calculate the upper and lower bounds of a 95%
> confidence interval for μ calibrates the population values so that their
> distribution is Normal.
> c. Given a 95% confidence interval for μ, the probability is 0.95 that μ is
> between the upper and lower bounds reported in the interval.
> d. Given a 95% confidence interval for μ, the probability is 0.95 that the
> mean, $\bar{x}$, of a new sample would fall between the upper and lower
> bounds reported in the interval.
> (MT2_Exam, question 8)

> How is a P-value to be interpreted?
> a. The P-value is the probability that H_0 is true.
> b. The P-value is the probability of a Type I error.
> c. The P-value is an assessment of the power of the test.
> d. The P-value measures the probability of observing patterns in the data
> at least extreme as what was observed if H_0 is true.
> (MT2_Exam, question 17)

Evidence abounds that Case B shares the GAISE goals related to student

understanding of the basic ideas of statistical inference.

Fifth block of goals. Table 11 lists the goals relate to critical thinking about

statistical results and presents the frequency of NVivo coding to each.

Table 11 Fifth Block of Goals, Coding Frequencies			
Finally, students should know…	Verbal	Written	Assessed
How to interpret statistical results in context	2	2	1
How to critique news stories and journal articles that include statistical information, including identifying what's missing in the presentation and the flaws in the studies or methods used to generate the information	1	0	0
When to call for help from a statistician	0	0	0
Note: Cumulative frequency does not match NVivo count because some items matched multiple goals. Frequency of Written occurances do not include the textbook.			

The contextualized interpretation of statistical results shares evidence with the previous two blocks. In the post-semester interview with Professor B, the question of ability to critique statistical reports elicited the following response:

> I think they would know how to talk to a colleague about some report that involves P values … As far as being effective citizens and consumers [pause] I'm not sure; I think so. When they hear a news story about some opinion polls and a margin of error, I think they would know how to interpret that.

Although the Professor is not observed to mention a time when it would be appropriate to consult a statistician, the textbook is explicit. "We have not discussed how to do inference about the mean of a clearly non-Normal distribution based on a small sample. If you face this problem, you should consult an expert" (Moore, et al., 2009, p. 440).

Overall, this block of goals has the weakest evidence of alignment between Case B and GAISE since two of the three goals have only a single, uncorroborated mention in the course.

Recommendations for teaching. The six recommendations for teaching presented in the GAISE College Report are on the observation protocol (see Appendix A)

with a four-point scale from "not present" to "major part." At the end of each observation the professor's inclusion of each recommendation received a rating. The Verbal column of Table 12 summarizes the frequency counts of observations where the recommendation was rated as having "part" or "major part" of the class. The coding of course documents through NVivo 9 provides frequency counts in the Written column of Table 12 from the lecture notes, recitation activities, and syllabus; the textbook is not available electronically for NVivo analysis. The counts in the Assessed column come from the two midterm exams, the cumulative final exam, and the twelve homework assignments.

Table 12 Recommendations for Teaching, Coding			
	Verbal	Written	Assessed
Emphasize statistical literacy and develop statistical thinking	5	5	1
Use real data	1	3	0
Stress conceptual understanding, not merely knowledge of procedures	6	8	8
Foster active learning in the classroom	2	0	0
Use technology for developing concepts and analyzing data	6	14	0
Use assessments to improve and evaluate student learning	0	1	0
Note: Frequency of Written occurances do not include the textbook.			

Emphasize statistical literacy and develop statistical thinking. Before the semester began, Professor B was asked what should be expected of students completing the course.

> They should be familiar with all of these techniques that I am talking about …
> when they graduate I don't expect that they would necessarily be able to perform
> all those procedures. I *would* expect them to be familiar with some things [like p-

83

value and confidence] … It's not so much about the technical ability but more about statistical literacy.

> Perhaps they should be able to identify when statistics are being misused. For instance, they could tell [if a sample] is clearly biased or haphazard. I would want the students to be able to recognize that; a red flag should go up. (Instructor interview, pre-semester)

When asked after the semester about the students' gain of statistical literacy, the response was both affirmative and decisive.

> They gained statistical literacy. They've definitely done that. I think that's one of the main objectives of the course. They know what a P value is and they know how to use it. They know what a confidence level is. I think they would know how to talk to a colleague about some report that involves P values.
> As far as being effective citizens and consumers [pause] I'm not sure; I think so. When they hear a news story about some opinion polls and a margin of error, I think they would know how to interpret that. (Instructor interview, post-semester)

The professor's concern for students' statistical literacy is evident in the first lecture. The initial discussion of statistics as "the science of data" (Moore, et al., 2009, p. 4) includes many references to everyday sources of statistical information, both descriptive and inferential (Lecture Topic Intro; Observation 1). A later observation shows interest in conveying the different uses for confidence intervals and significance tests as well as a layman's interpretation of p-value: "The smaller p-value indicates a more surprising pattern" (Observation 3).

The final exam includes one question that is purely an assessment of the students' statistical literacy:

> Suppose the correlation between variables x and y is of a magnitude near one (i.e., $|r| \approx 1$). What does this indicate?
> > a. The phenomenon measured in x causes that measured in y.
> > b. The phenomenon measured in y causes that measured in x.
> > c. There may or may not be a causal relationship between phenomena measured in x and y.
> > d. Both "a" and "b".

Several questions on the two midterm exams and some homework questions straddle the line between literacy and conceptual understanding. These were coded in the other category (Stress conceptual understanding rather than mere knowledge of procedures) and are discussed below.

Development of statistical thinking is not explicitly evident in Case B. GAISE describes statistical thinking as "understanding the need for data, the importance of data production, the omnipresence of variability, and the quantification and explanation of variability" (ASA, 2005, p. 14). The evidence presented for the first block of goals for students may also be applicable to this teaching recommendation even though the Case does not use the term "statistical thinking" anywhere.

Use real data. The first observation on the first day of class included live web links to several public sources of data. Professor B speculated on how each might be useful for answering business and personal questions (Observation 1). A technical report on a health study brought some basic ideas of inference to the attention of students (Observation 1; Sanders, 2011). These are the most explicit instances of evidence that Case B strives to help "students learn to formulate good questions and use data to answer them appropriately" (ASA, 2005, p. 16).

"A lot of the lecture data came from the textbook. The textbook claims to have real data," Professor B said in the post-semester interview when asked about the examples used in lectures. In fact, the textbook includes 24 "cases" across the twelve chapters covered by Case B (e.g, Uncovering Fraud by Digital Analysis, Moore, et al., 2009, p. 249). Each of these cases provides an endnote citation if not the actual data from which the summary statistics or inferential conclusions were drawn. Some examples in

the chapters also include endnote citations. When cases or examples from the textbook are presented in lectures (e.g., Predicting College GPA, Moore, et al., 2009, p. 638 and Lecture Topic 12; see Appendix C for a Topic 11 example), the citations are lost. Students are unaware that the data they see is real even when it is, unless they are reading the textbook closely and checking the endnotes.

Chapter exercises sometimes refer to the data used in cases or examples. Exercise 11.59 (p. 653) refers to the data used in the Predicting College GPA example. The use of web-based homework rather than the textbook exercises, however, makes it likely that only the most diligent students would ever take the opportunity to work with real data themselves. The data that students may take time to investigate comes in the Excel Demos that are used in lectures and available to them through the course management system, though there is no requirement for them to do so. Even if they did, Professor B admits, "The Excel data used for demonstrations was mostly invented. There was no context" (Instructor interview, post-semester).

Stress conceptual understanding rather than mere knowledge of procedures. At the beginning of the semester, Professor B remarks on the course objectives: "To encourage a sense of critical thinking and questioning … I try to be very clear about what the concepts are in my lectures" (Instructor interview, pre-semester). Asked about students who are successful in the course, the professor returns to the need for understanding concepts:

> My sense of the students who succeeded were the ones thinking about the concepts. Some students didn't want to put the effort in; they wanted to figure out how to rely on their calculators to do the numbers. It's kind of a different attitude about the subject. Some students just want to know what numbers to put in the calculator; some students are trying to think and put things together. So I think

the students that were putting in the effort to put things together … I just get the sense that they were more successful. (Instructor interview, post-semester)

The professor echoes the GAISE attitude about conceptual understanding: "If students don't understand the important concepts, there's little value in knowing a set of procedures. If they do understand the concepts well, then particular procedures will be easy to learn" (ASA, 2005, p. 17).

The early lectures on inference also demonstrate the instructor's interest in the conceptual understanding of students (Lecture Topic 6; Observation 3). "Probability calculations help distinguish patterns seen in data between those that are due to chance and those that reflect a real feature of the phenomenon under study … We're rethinking the probability when things are no longer random since we've observed the variable" (Observation 3). The first part of that statement referred to earlier lectures that mentioned statistical significance as the outcome of rigorously designed studies; the latter half references the difference between significance testing and estimating with confidence.

A recitation activity for Topic 8 provides another opportunity to discuss the connections mentioned in lecture— "More confidence → margin of error increases → wider interval → less precision" (Observation 3)—by comparing intervals calculated by hand (estimating degrees of freedom) and those calculated by computer. "Notice that the conservative estimate is wider than the confidence interval created by the software using the better estimate of the degrees of freedom" (Recitation 1). A similar comparison activity asks about the use of the Standard Normal (z) or Student's t distributions. Professor B admits, "We'd almost never see this in practice but it is good to think about for conceptual understanding" (Recitation 1).

Exam questions are conceptual in nature for about one-third of the questions (33.33% collectively, with each exam ranging between 27-37% individually). The homework and recitation exercises disappointed Professor B: "The TAs tended to pick out questions [for recitations] that take what was learned in class and show how it could go in interesting directions" and "[homework] questions again are focusing on the sticking points. I'm not sure that it's really focusing on the important concepts. More of 'do you know the rule?' instead of 'do you know the concept?' … In the future, I think I would select questions from the textbook" (Instructor interview, post-semester).

Foster active learning in the classroom. Professor B says, "I have looked for places to do more discussions … looking for students to express their ideas verbally and explain what they are thinking and not relying on the mathematics. I do try to do that, but there are very few places" (Instructor interview, pre-semester). Later in the interview, the instructor spoke about the most likely opportunity for students to have discussion is in the recitations: "I've asked the TAs to divide the students into small groups. I haven't done that before but that is, again, to encourage discussion between students." This adjustment to the previous semester's use of the recitation sessions is a step toward the GAISE vision for active learning through "group or individual problem solving, activities and discussion" (ASA, 2005, p. 18).

On the two occasions counted in Table 12, Professor B spoke to students about the importance of talking about their work. "I like when students work together because when you have to talk about it you seem to learn it better. When you have to explain it to someone else it helps you to learn that concept" (Observation 1). During the first observed recitation, the idea was reiterated: "Talk to your neighbor. It's good to discuss

your ideas with a neighbor. When you have to explain them to someone else it is helping you to learn, so I encourage you to do that."

One exceptional instance of active learning took place in the first lecture when students are asked to contribute to a comprehensive definition of "statistics." Professor B provided definitions from a number of places and exhibited websites with data as well as a completed study report as a pre-cursor to the activity. Participation ranged across the auditorium with both men and women making suggestions. In other observations, open questions received fewer students responses.

Use technology for developing conceptual understanding and analyzing data. Every topic has an accompanying Excel spreadsheet that precisely meets the GAISE suggestion to "perform simulations to illustrate abstract concepts" (ASA, 2005, p. 20). Although not used in every observed class period, some topics took more than one lecture period to complete and Professor B did remark to that class that the "Excel demonstrations—each week I give you one—are not always useful to every student but please try this one" (Observation 4). They are all available through the course management system for students to investigate for themselves, another of the GAISE suggestions.

The spreadsheet for Topic 7 (see Appendix D) is an example of additional alignment with GAISE suggestions regarding interactive capabilities and dynamic linking between data, graphical, and numerical analyses. It includes a population of 100 values and 100 random samples of size 20 with their respective means (this much was also in the example for Topic 3 to demonstrate sampling distributions). P-values for a two –sided significance test as well as a confidence interval for each are also calculated. Additional

cells indicate rejection of H_0 and intervals that do not include the parameter, both linked to graphical displays. The professor demonstrates changing significance/confidence values, and encourages the students to try this themselves (see quote in previous paragraph).

Professor B sometimes uses Excel to calculate p-values for example problems during lectures (e.g., Observation 5). By the time the course reaches multiple regression, all calculations are done by software. The textbook includes output from Excel, SPSS, Minitab, and SAS for the example that is used in the lecture (Moore, et al., 2009, p. 641-642; Lecture Topic 12).

Use assessments to improve and evaluate student learning. The online homework system provides the most feedback to students of any course assessment. In the first recitation meeting of the semester, Professor B goes over the use of the system. Some problems allow for multiple attempts, some allow partial credit, and only the last attempt is graded at the due date. Feedback about likely error is not available in this system but correct answers are available after the set is graded.

Exams are all multiple-choice and graded by optical mark recognition (i.e., Scantron), a necessity in such a large class. The exam questions and answer key are made available in the course management system, while the students' individual answers are recorded in the comment attached to their score in the grade book. Like the homework, this does provide some feedback but is not necessarily useful for improving students' learning. Explanations to accompany the answer key might be a step closer to GAISE's assertion that "useful and timely feedback is essential for assessments to lead to learning" (ASA, 2005, p. 21).

Summary. Case B demonstrates alignment of its course objectives with GAISE Goals for Students. The evidence is strongest in the area of understanding the basic ideas of inference, with triangulation among the verbal, written, and assessed categories of evidence. Goals in the block about statistical processes for answering questions also match Case B objectives, though the evidence is somewhat less plentiful and not perfectly triangulated. The remaining blocks have uneven evidence of alignment with GAISE.

The use of technology for developing concepts and analyzing data in Case B is consistent with GAISE recommendations for teaching. Stressing conceptual understanding and emphasis on statistical literacy is also evident in Case B. The other three recommendations—active learning, real data, and assessment for learning—have little evidence of alignment with GAISE.

Case C – Statistics for Students in a Social Science Major

The setting. The students required to take this course share a common major (or minor) in a social science discipline. There are no pre-requisites for taking this course; it is the pre-requisite for the research methods course. Students generally take both courses in their sophomore year, but students at other points in their undergraduate career are not uncommon (Instructor interview, pre-semester). One of the first activities done in class is a review of selected arithmetic topics including order of operations, fractions, negative numbers, square roots, and basic algebra (MathAssessment).

There are a total of 33 students enrolled in two sections. All students meet at the same time for lectures but there are two groups for lab meetings. Observations were done on different days of the week to capture lectures before and after the labs as well as with

and without weekly quizzes. The classroom is furnished with columns of individual desks, a computer desk for the instructor, a projector, and whiteboards at the front which are covered by the screen when the projector is in use. There are more women than men enrolled in the course (24:9).

Professor C has been teaching this course for thirteen consecutive semesters, the entirety of the professor's affiliation with the institution. "Nobody fights me to teach this class" (Instructor interview, pre-semester). The professor earned bachelors, masters, and doctoral degrees in this discipline from the mid-1990s to the mid-2000s. A postdoctoral fellowship at a different institution preceded the current appointment as assistant professor (Instructor CV).

Course design. An online course management system is available to course participants, including the observer. The resources provided through this system are the presentation slides for the lectures (see examples in Appendix C), data files used during lab sessions, the syllabus, and details for the research assignment. There are also links to the textbook companion site and the online homework system. Grades are entered regularly in the gradebook as the course progresses.

Lectures follow the textbook's order of presentation with little deviation. The syllabus indicates the predetermination to skip the section on multiple regression, the chapter on the binomial distribution, and parts of chapters on hypothesis testing. Delays in the textbook delivery put the actual lectures off the published schedule early in the semester and never fully recovered (Instructor interview, post-semester). Analysis of variance received less coverage than planned and two-way ANOVA only received brief coverage in the last lecture (Observation 7). The semester ended before any

nonparametric tests could be addressed in lectures or labs, in spite of their appearance on the course schedule.

The course has a dedicated tutor who offers study sessions at least once a week. The tutor (a senior majoring in the discipline) attends lectures but does not participate in activities. On one occasion the tutor led the class by returning and reviewing an exam because the professor was conducting an experiment with another class that met at the same time.

Assessments. A variety of assessment tools are used in the course. These include lab assignments with a culminating binder, online homework assignments, weekly quizzes, a research proposal, four exams, and a comprehensive final exam. Lab assignments often require students to use the statistical software SPSS and, collectively, the assignment instructions provide a good resource for further use of SPSS in the research methods course (Instructor interview, pre-semester). The research proposal requires students to review some discipline literature to provide background for an experiment they would like to conduct in order to test an original hypothesis. They specify the variables, participants, the actions participants will take, and the statistical test appropriate to the data they (hypothetically) collect and the question they hope to answer (ResearchAssignment). "Within that, I want to make sure that in the design of their study they picked the right test. That's kind of the big piece… That paper is sort of like the essay portion of the final exam" (Instructor interview, pre-semester).

The online homework system available through the textbook publisher is a new course feature this semester. There were problems getting the textbooks, which had been ordered directly from the publisher as a bundle with the access code for the homework

system. "A number of students had their orders canceled because there was an error on the website that packaged a year of access to the online homework with the eBook instead of just one semester" (Instructor interview, post-semester). Some students prefer the eBook, but all students need the hardcopy of the textbook in order to keep their access to the homework system (Observation 2).

Homework, quiz, and exam questions are a mix of definitions, computations, and interpretations. For example, Quiz 2 asks students to "identify the scaling of the following variables" and "identify whether the following variables are discrete or continuous" as evidence that they understand the definitions of variable characteristics. The fifth homework assignment presents a series of computations based on a normally distributed variable including "You can infer that 97.72% of the female students have scores above _____." Exam 3 contains several examples of interpretive questions such as "Is it correct to conclude by 'accepting' H_0 when the results of an experiment are not significant? Explain."

Exams include many multiple-choice questions that assess student understanding of basic concepts and definitions. These are supplemented by open response questions such as the one quoted in the previous paragraph that are also focused on conceptual understanding. Together these represent 51-74% of each exam. A small portion of the remaining exam questions blend computation with interpretation, representing a mix of conceptual and procedural knowledge. Strictly procedural questions are most prevalent in the first and last exams (44% and 36%, respectively, of exam content).

Case C – Statistics for Students in Social Science Majors, Pattern Matching Analysis

Like the previous instructors, Professor C had no previous knowledge of GAISE (Instructor interview, post-semester). Once again the lack of awareness does not preclude mutual goals for students or implementation of recommended pedagogy.

Goals for students. The observation protocol (see Appendix A) was used to tally the instructor's verbal remarks concerning topics listed in the five blocks of goals for students in the GAISE College Report (ASA, 2005). NVivo 9 coding of interview transcripts and observation notes provides frequency counts of Verbal evidence that are reported in the tables that accompany this analysis. Additional coding of lecture notes and the syllabus provided the frequencies of Written evidence. Coding of exams, lab assignments, quizzes, and the research project provides evidence of what goals were Assessed during the course. Homework assignments and the textbook were not available electronically for NVivo coding; therefore, some evidence is not included in the tables' counts but still contributes to the analysis that follows each table.

This quantitative estimate demonstrates that Case C's goals for the students align with most of the items in the five blocks of the GAISE list (see Tables 13 through 17).

First block of goals. Recall that the goals in this block (see Table 13) relate to important concepts about what information statistical analysis can and cannot provide.

Table 13
First Block of Goals, Coding Frequencies

Students should believe and understand why…	Verbal	Written	Assessed
Data beat anecdotes	1	2	1
Variability is natural, predictable, and quantifiable	1	0	0
Random sampling allows results of surveys and experiments to be extended to the population from which the sample was taken	0	2	0
Random assignment in comparative experiments allows cause-and-effect conclusions to be drawn	0	1	0
Association is not causation	0	1	1
Statistical significance does not necessarily imply practical importance, especially for studies with large sample sizes	1	1	0
Finding no statistically significant difference or relationship does not necessarily mean there is no difference or no relationship in the population, especially for studies with small sample sizes	0	0	0

Note: Cumulative frequency does not match NVivo count because some items matched multiple goals. Also, the Written and Assessed are undercounted because the textbook, homework assignments, and the final exam were not available electronically.

Case C is the only one of the four cases in this study to explicitly address student beliefs about the benefit of data over other methods of knowing. Chapter 1 of the textbook and the first lecture address four methods of knowing: authority, rationalism, intuition, and scientific (Lecture1; Observation 1; Pagano, 2010). The scientific method is described as beginning with a hypothesis that comes from one of the other methods but "data from the experiment force a conclusion consonant with reality" (Pagano, 2010, p. 6). The topic is important enough to the instructor that student understanding of the distinctions between the four methods is assessed with two open-ended homework questions (ProblemSet_Chapter1) and a multiple choice question in the first exam.

Variability is not treated explicitly in lecture notes or in any observed classes. The textbook introduces measures of variability as a means to "quantify the extent of dispersion" (Pagano, 2010, p. 79). At the beginning of the chapter on measures of central tendency and variability, the importance of this quantification is stated as "the need to know whether the effect of the program is uniform or varies over the youngsters. If it varies, *as it almost assuredly will,* how large is the variability?" (Pagano, 2010, p. 70, emphasis added). Professor C does mention a lab assignment that addresses variability, "I get a box of Skittles and we weigh all of the bags to look at the variation. The students are usually surprised that they don't all weigh the same" (Instructor interview, pre-semester).

The need for random sampling is introduced in the first lecture as a characteristic of "true experiments … Random sampling increases the chance that the sample will mirror the population" (Lecture1, p. 23). There is no further discussion of its importance until Chapter 8 in the textbook, when random sampling and probability are introduced as crucial to meaningful inference (Lecture13; Pagano, 2010). Random assignment is mentioned in both the textbook and the lecture notes but receives minimal emphasis, although most of the examples of inference are experimental rather than observational.

Lecture 9 addresses the difference between causation and association somewhat obliquely: "Sometimes we cannot run an experiment to determine cause and effect. Instead, relations can be determined between two variables." The textbook is more detailed in its discussion of the implications of correlated variables. A multiple choice question on the second exam assesses student understanding of the concept:

> Knowing nothing more than that IQ and memory scores are correlated 0.84, you could validly conclude that ____.

a.   good memory causes high IQ
b.   high IQ causes good memory
c.   neither good memory nor high IQ cause each other
d.   a third variable causes both good memory and high IQ
e.   none of the above

The distinction between statistically significant results and practical importance appears briefly in the lecture that introduces hypothesis testing. "Are the results important? Effect may be significant but small" (Lecture15). However, the impact of sample size is not discussed and student understanding is not assessed. The textbook includes a section titled "Size of Effects: Significant Versus Important" (Pagano, 2010, p. 256) that mentions that a large sample may detect a small effect. More details about effect sizes and the relationship to statistical significance are covered in the textbook chapter on Analysis of Variance but it does not appear in the lecture materials for that topic.

While the final goal in this block did not receive any coding, part of this understanding is implied in the lecture about power. "Power varies directly with the size of the real effect of the independent variable" (Lecture17, slide 2) and may be calculated "when our experiment failed to reject the null hypothesis" (Lecture17, slide 4). Professor C elaborates, "Retaining the null may mean that we didn't have enough power to detect the change. We may do the experiment over again, perhaps with a larger sample" (Observation 5).

Case C provides evidence that this course holds some of the same goals for students' beliefs and understanding as listed in GAISE. Particular care is shown for describing the need for data over other methods of knowing. Other concepts in this block are mentioned in lectures and receive additional attention in the textbook.

Second block of goals. Table 14 lists the goals in this block that relate to appropriate interpretation of results from statistical analyses.

Table 14 Second Block of Goals, Coding Frequencies			
Students should recognize…	Verbal	Written	Assessed
Common sources of bias in surveys and experiments	0	0	1
How to determine the population to which the results of statistical inference can be extended, if any, based on how the data were collected	0	0	0
How to determine when a cause-and-effect inference can be drawn from an association based on how the data were collected (e.g., the design of the study)	0	1	1
That words such as "normal," "random," and "correlation" have specific meanings in statistics that may differ from common usage	0	0	0
Note: Cumulative frequency does not match NVivo count because some items matched multiple goals. Also, the Written and Assessed are undercounted because the textbook, homework assignments, and the final exam were not available electronically.			

Recognizing common sources of bias receives minimal treatment in Case C. The lecture introducing probability mentions the importance of random sampling in order to use rules of probability for making inferences from sample information to populations (Lecture13; Pagano, 2010). The textbook illustrates the concept with the example of the drastically inaccurate prediction of the 1936 presidential election due to biased sample selection (Pagano, 2010, p. 181). There is no evidence that this example is repeated during the unobserved lecture. The lecture on sampling distributions revisits the idea of a representative sample as important to making inference (Observation 5) and the first exam includes a multiple choice question on the subject. No other sources of bias appear in the text or lecture notes. In reviewing this analysis with the participating instructor,

Prof C indicated that there is a verbal conversation about selection bias accompanying the lecture on random selection (Personal communication, post-analysis).

The first chapter of the textbook and the accompanying lecture defines a sample as a subset of the population under study and further describes inferential statistics as techniques that allow sample data to be used for drawing conclusions about populations (Lecture1; Pagano, 2010). The same sources discuss "true experiments" as the only way to determine a cause-and-effect relationship. The first homework assignment presents the design of three studies and asks students to identify them as an observational study or a true experiment (ProblemSet_Chapter1). Student understanding of the distinction between observational and experimental studies is assessed with an open-ended question on the first exam: "How does natural observation research differ from true experiments?"

Distinction between statistical and every day usage of language is not addressed in Case C. GAISE provides three words as examples that may cause confusion during an introductory course: "normal," "random," and "correlation." There are no explicit definitions of the first two words in either the text or lecture notes; not even as statistical terms that might elicit student notice of the different usages. The definition of correlation is not contrasted with an everyday usage (Lecture9; Pagano, 2010).

Lack of emphasis on the topics GAISE suggests that students should recognize leaves this block with the least evidence of mutual goals.

Third block of goals. Procedures for obtaining and analyzing data with appropriate techniques and meaningful communication of the results comprise this block of goals (see Table 15).

Table 15 Third Block of Goals, Coding Frequencies			
Students should understand the parts of the process through which statistics works to answer questions…	Verbal	Written	Assessed
How to obtain or generate data	0	0	0
How to graph the data as a first step in analyzing data, and how to know when that's enough to answer the question of interest	4	1	1
How to interpret numerical summaries and graphical displays of data—both to answer questions and to check conditions (to use statistical procedures correctly)	1	3	1
How to make appropriate use of statistical inference	0	9	3
How to communicate the results of a statistical analysis	1	0	1
Note: Cumulative frequency does not match NVivo count because some items matched multiple goals. Also, the Written and Assessed are undercounted because the textbook, homework assignments, and the final exam were not available electronically.			

Case C makes no reference to *how* data is obtained or generated other than the need for random sampling technique as already discussed in relation to the first two blocks of goals. Students do participate in data collection through lab assignments (Lab_1, Lab_3) that possibly leads to informal discussion of various difficulties and careful techniques. These labs were not observed and later use of the collected data did not refer back to such conversations.

Consecutive chapters of the textbook outline procedures for construction of graphs and calculation of numerical summaries (Pagano, 2010). Both chapters mention indicators of distribution shape but do not reference them again in connection with checking conditions before testing. Professor C does reference descriptives as preliminary analysis: "Charts are often good ways to get a quick sense of what kinds patterns are in the data" (Observation 2).

An activity utilizing a classroom response system (clickers) took place at the beginning of Lecture 6 as a formative assessment of student understanding of the appropriate use of a bar graph versus a histogram. The online homework sets for these two chapters are mainly procedural types of questions (e.g., completing frequency tables, calculating percentiles, calculating the arithmetic mean or standard deviation) with a couple of interpretive or conceptual questions in each set (ProblemSet_Chapter3; ProblemSet_Chapter4). Examples of the non-procedural types of questions come from Exam 1, which gives equal weight to procedural/computation questions:

> Let's assume that we have determined the salary of all the professors at your school. In plotting the distribution of salaries we notice that it is positively skewed. For this distribution ____.
>   a. mean = median
>   b. mean < median
>   c. mean > median
>   d. can't tell from the information given

> When might the median be a better statistic to use for central tendency than the mean? Illustrate your answer by using an example.

When asked about the student outcomes, Professor C replied, "I think the most important thing is to understand what kinds of analyses are appropriate to use when, and why that is" (Instructor interview, pre-semester). Quiz 12 includes three testing scenarios for which students must name the type of test appropriate for analyzing the collected data. Conducting the proposed *t* tests takes up the rest of the class period (Observation 6).

One student works alone rather than with his group on the practice t tests and finds the first p-value very quickly by using his calculator. He asks for the next data set but Professor C will not let him move forward without completing the intermediate calculations and, most especially, stating a conclusion in terms of the context

(Observation 6). Lab assignments invariably ask for final statements to connect the statistical result with the original context. For example, students finish a lab by answering a question like this: "If $\alpha = 0.05$, do you retain or reject the null? Give an interpretation of these results." (Lab7).

Case C contains many indicators that students learn procedures for descriptive statistics, including graphs and exploratory data analysis. Connecting these things to checking conditions for testing is not evident but there is emphasis on the selection of appropriate inferential procedures and effective communication of results.

Fourth block of goals. The goals in this block (see Table 16) relate to important concepts needed for accurate interpretation of inferential analysis.

Table 16 Fourth Block of Goals, Coding Frequencies			
Students should understand the basic ideas of statistical inference…	Verbal	Written	Assessed
The concept of a sampling distribution and how it applies to making statistical inferences based on samples of data (including the idea of standard error)	2	5	2
The concept of statistical significance, including significance levels and p-values	4	5	3
The concept of confidence interval, including the interpretation of confidence level and margin of error	0	0	0
Note: Cumulative frequency does not match NVivo count because some items matched multiple goals. Also, the Written and Assessed are undercounted because the textbook, homework assignments, and the final exam were not available electronically.			

During the pre-semester interview, Professor C identified sampling distributions as a topic that is especially troublesome for students:

> The fact that we've up to this point been talking about individual scores: where individual scores fall within the sample or population. Now we have to start thinking about samples…You can no longer take your sample and compare it to an individual. It's not fair to compare 30 people to one person. So it's only fair to take those 30 people and compare it to all other possible groups of 30 people. We start with the raw score distribution, then we look at sample distribution. We talk

about how the means are the same but the standard deviations are different. They *cannot* wrap their heads around it.

Lecture 17 introduces the concept of sampling distributions, Lecture 18 is devoted to the development of the idea, and Lecture 19 begins with a review. Additionally, Professor C presents a spreadsheet that contains multiple samples from a population of values that students used in a previous lab. "You can see that most of these sample means hover around the population mean of 7. Similarly, the standard deviations, in general, hover around 3.74" (Observation 5). The textbook devotes a chapter to the topic and includes a figure illustrating all possible samples of size two from a population of five scores (Pagano, 2010, p. 289-93). There is an accompanying homework set that includes multiple-choice questions about the characteristics of sampling distributions as well two opportunities to calculate the standard deviation of the sampling distribution of sample means (ProblemSet_Chapter12). Exam 3 includes one multiple-choice and one open-response question about the conceptual construction of a sampling distribution.

Significance levels and p-values are inextricably linked to all of the hypothesis tests covered by the course. As such, they receive some attention in every lecture and textbook chapter after their introduction in Lecture 14 and Chapter 10 (out of 24 lectures and 18 chapters). Lecture 15 includes a discussion of Type I and Type II errors that is prefaced with the following:

> Why $\alpha = 0.05$? This means that 5 out of 100 times the result could lead you to reject the null when it is actually true. If the experiment is replicated (do the experiment again, you or another researcher), they may continue to get null results and you will feel like you have "egg on your face" even though you did nothing wrong. You could lower the value of α to avoid that feeling but it comes at a cost. (Observation 4)

The homework set for Chapter 10 includes some fill-the-blank questions about the use of α, and lab assignments 8 through 11 include identification of the critical value. Lab 10 has a multi-step exercise that leads students to see the connection (through α) between decisions based on critical value or p-value. There are four multiple-choice questions about the meaning of α or statistical significance on Exam 3.

The transition from probability topics to hypothesis testing alludes to a familiar definition of p-value: "Why do we need to know about probability? In inferential statistics, [we need to know the] probability of getting our obtained result or something more extreme by chance" (Lecture14). Professor C uses similar wording when discussing the decision to reject or not reject a null hypothesis by saying, "The p-value is equal to the probability that getting a test value this far from the mean – or even farther – might happen by chance" (Observation 4).

Students get their first opportunity to determine p-value and make a decision about significance in a lab activity where they are asked, "Is your weight for sample 1 unexpected? In other words, is sample 1 significantly heavier or lighter than expected (note the p value)?" (Lab8). After providing information about an experiment, Exam 3 asks students to do the following:

> Calculate the appropriate test statistic and make a conclusion based on your result. Note both your test value and either the p value or critical value for your test. (Assume population normality and use $\alpha = 0.05_{1 \text{ tail}}$.)

Exam 4, however, presents four separate tests that specifically ask for the critical value immediately preceding the question of inference, "What do you conclude, using $\alpha = 0.01_{1 \text{ tail}}$?"

There is no evidence of instruction regarding confidence intervals. The textbook contains a four-page introduction to confidence intervals for estimating the mean (Pagano, 2010, p. 331-334). There are no accompanying lecture notes or assessment items. In the post-analysis review, Professor C explained the missing material as part of the schedule adjustment.

Case C demonstrates heavy emphasis on sampling distributions and statistical significance as key components of statistical inference. The course is in accord with the fourth block of goals for students outlined in GAISE with the exception of confidence intervals.

Fifth block of goals. Finally, this block of goals (see Table 17) relate to critical thinking about statistical results.

Table 17 Fifth Block of Goals, Coding Frequencies			
Finally, students should know…	Verbal	Written	Assessed
How to interpret statistical results in context	2	1	2
How to critique news stories and journal articles that include statistical information, including identifying what's missing in the presentation and the flaws in the studies or methods used to generate the information	0	2	0
When to call for help from a statistician	0	0	1
Note: Cumulative frequency does not match NVivo count because some items matched multiple goals. Also, the Written and Assessed are undercounted because the textbook, homework assignments, and the final exam were not available electronically.			

Evidence from Case C for the goal of students being able to interpret statistical results in context is already presented in blocks one, two, and three. The emphases on statistical significance versus practical importance, non-significance versus no difference,

determination of appropriate population for inference, and communicating results of statistical analysis all contribute to the goal as listed in this block.

In the first interview Professor C addressed the critique of news stories in two separate contexts. The discussion of students' preparation for taking the class led to a mention of a study about math anxiety that is presented to students as both intervention for their personal anxiety and an example of critiquing published statistics. "We'll talk about this today, about why it's important to know something about statistics. Being able to look at an article like this and try to figure out … can they really make the conclusions that they are making?" Later in the interview the discussion of what students should gain from the course returns to this topic: "Papers don't always get it right, or they overstep a lot in terms of the conclusions that they draw about a lot of things. Hopefully this class teaches them to be good consumers about what they read, to think critically about what they see." The textbook includes a dozen critiques of published statistics in sections labeled "What is the Truth?" (Pagano, 2010).

Journal articles get separate attention during class time. In Observation 4 the consequences of non-significant results include the professor's assertion that "little is published when the null hypothesis is supported by the experiment; there are *some* 'no difference' results that are interesting" (emphasis in original). Additional attention to journal articles occurs in the next observation:

> This class is important because you may be in a position to review other [scientists'] work in addition to being able to use statistics in your own research. Without the researcher's data, we can only go on what they report in their methodology section to know if they had a good design and used the right test. (Observation 5)

The last goal for students receives one explicit mention, though it is not the target

idea being assessed with the following question from Exam 2:

A correlation between college entrance exam grades and scholastic achievement
was found to be -1.08. On the basis of this you would tell the university that ____.
a. the entrance exam is a good predictor of success
b. *they should hire a new statistician* [emphasis added]
c. the exam is a poor predictor of success
d. students who do best on this exam will make the worst students
e. students are this school are underachieving

The evidence from Case C shows that it shares the GAISE goals for students

listed in this block.

Recommendations for teaching. The observation protocol (see Appendix A)

once again provided the frequency counts for the Verbal column in Table 18, while

frequency counts from NVivo 9 provided initial analysis on Written (lecture notes, in-

class activities, and the syllabus) and Assessed (quizzes, project, exams) documents.

Table 18 Recommendations for Teaching, Coding Frequencies			
	Verbal	Written	Assessed
Emphasize statistical literacy and develop statistical thinking	4	14	14
Use real data	2	0	3
Stress conceptual understanding, not merely knowledge of procedures	5	25	10
Foster active learning in the classroom	5	10	11
Use technology for developing concepts and analyzing data	0	4	11
Use assessments to improve and evaluate student learning	4	0	2
Note: Frequency of Written occurances do not include the textbook; Assessed counts do not include homework or the final exam.			

Emphasize statistical literacy and develop statistical thinking. Case C presents strong evidence that statistical literacy and thinking are important objectives of the course. Understanding the language and fundamental ideas of statistics (ASA, 2005) are at the heart of the instructor's expectations for the students: "Hopefully this class teaches them to be good consumers about what they read, to think critically about what they see" (Instructor interview, pre-semester). The textbook includes a dozen *What is the Truth?* articles that critique published studies, advertising, or popular media reports of research (Pagano, 2010). Although these are not specifically assigned for student reading and not mentioned in lectures, they are an available resource to students that models both statistical literacy and thinking. One very appropriate assessment of statistical literacy that appears on the first exam is the open response question "When might the median be a better statistic to use for central tendency than the mean? Illustrate your answer by using an example."

Professor C also encourages student interest in gaining statistical literacy and developing statistical thinking by connecting the course to their discipline: "This class is important because you may be in a position to review other [scientist's] work in addition to being able to use statistics in our own research" (Observation 5). The last statement was made in part of a conversation about a researcher's faulty publications that made a national newspaper in the previous week. The discussion of Type I and Type II errors led to a conversation about false imprisonment based on eyewitness accounts. Professor C reminded the students of research results presented to them the previous year, during the introductory course to the discipline (Observation 4).

As suggested by GAISE, students were given opportunities to choose appropriate techniques for graphing (Lecture6) and hypothesis testing (Lecture23), not merely implement a task imposed by the instructor. Also in line with GAISE suggestions, students in this course engage in an open-ended project where they must use statistical thinking to design a study for answering a question of their own (Research_Assignment). This culminating activity agrees with Professor C's expectation for what students should know and be able to do by the end of the course: "I think the most important thing is to understand what kinds of analyses are appropriate to use when and why that is… just how to be good consumers. To try to teach them that skepticism…" (Instructor interview, pre-semester).

   Use real data. The first day of the semester was also a lab day for half of the class (the other half went the next day) and the first lab assignment was a 47 question survey. This task illustrates Professor C's focus on using real data in multiple ways, as suggested by GAISE. Answering the survey introduces students to SPSS by careful consideration of the variables and entry of their own data (Lab_1). The instructor merged the class files so that a larger data set was available for later analyses. The height data informed the construction of frequency distributions (Observation 2), the relationship between high school GPA and the number of extracurricular activities was explored through regression (Lecture12) and the introduction of *t* tests for independent samples includes team practice using the extracurricular activity data (Lecture21). Another data collection activity required weighing bags of Skittles followed by practice with descriptive statistics (Lab3). This data set is revisited with the introduction of z scores (Instructor interview, pre-semester) and sampling distributions (Lab_8).

The homework sets include data from published sources (e.g., the *Brown Corpus of Standard American English* in Problem Set 2, and the U.S. Census Bureau in Problem Set 12). There are also sets of data that are more likely to be only realistic that describe research typical in the discipline. For example, the following scenario is presented ahead of calculations leading to a correlation coefficient:

A [researcher] is studying trends in childbearing. She asks expectant parents in different parts of the country about the number of children in their family of origin and the total number of children they plan to have. On an average day of data collection, she gets the following results:

Parent	Children in Family of Origin (X)	Number of Planned Children (Y)
1	3	2
2	2	3
3	1	0
4	3	2
5	4	5

Similar scenarios are also in some lab assignments (e.g., Lab_10) in addition to the use of the data collected from the students (e.g., Lab_9).

Stress conceptual understanding rather than mere knowledge of procedures.
Professor C shows interest in the conceptual understanding of the students but is also aware of the need for practicing procedures. "We talk about the concepts in class but mostly when I [create] homework it is doing problems, practicing doing problems because they're afraid of the math. What I liked about [using online] homework is that it gives them some conceptual questions as well" (Instructor Interview, post-semester). Every exam included conceptual questions that continued the assessment of student understanding beyond mere knowledge of procedures, such as: "Which test, z or t, has higher power? Explain why" (Exam4).

Observation 4 included some interesting discussion about the related concepts of

significance, p-value and decision errors.

> Why α = 0.05? This means that 5 out of 100 times the result could lead you to
> reject the null when it is actually true. If the experiment is replicated (do the
> experiment again, you or another researcher), they may continue to get null results
> and you will feel like you have "egg on your face" even though you did nothing
> wrong. You could lower the value of α to avoid that feeling but it comes at a cost.

After the introduction and definitions of Type I and Type II errors, the Professor connects

back to the previous discussion of the level of significance:

> "Which is worse?"
> (Polled class; a few students think Type II is worse, some abstain)
> "Who feels the pain of a Type II error?"
> The researcher does – replicate with improved power; it could withhold useful
> treatments. If you really believe in the effect of your treatment, you probably re-
> run the experiment to see if you get the same results.
> "Who feels the pain of a Type I error?"
> The public – worst case is putting out advice/products that actually cause harm (a
> student suggests Vioxx). The researcher might be embarrassed to find that others
> cannot replicate the results.

In between these conversations is a slide entitled "Are research findings always the

Truth?" (Lecture15). A few students are quick to answer "no" and the professor agrees.

"When we publish results we cannot say that we have 'proven x' but that 'almost beyond

a reasonable doubt,' we think this happens or it supports the [alternative] hypothesis. We

can actually never know the true reality" (Observation 4). The usual error table for

discussion of the two types follows this. After the discussion of which is worse, the

professor adds, "It's like putting your data on trial" and the next slide includes famous

cases of judicial error that prompt the previously mentioned connection to false

imprisonment and the students' prior exposure to relevant research.

The syllabus for Case C includes a schedule organized by statistical techniques

rather than focused on key concepts as suggested by GAISE. The intended breadth of the

course, however, did not take place. In the effort to schedule observations of particular topics, Professor C responded with details about the adjustments to the original schedule: "I have been using lab to try to get back on schedule ... I think that I have probably used about 3/4 of a class period because of textbook issues" (Personal communication, week 3). Three weeks later, during an observation of a lab session, lecture material preceded the assignment (Observation 3), which indicated that the impact of the textbook issue was not the only factor influencing schedule changes.

In the end, Analysis of Variance received minimal treatment—two classes instead of the five planned—while the three classes on non-parametric tests did not happen at all (Observation 6). In contrast, sampling distributions took three classes (Lecture17, Lecture18, Lecture19) instead of the two on the schedule and correlation topics extended across three classes (Lecture9, Lecture10, Lecture11) plus the introduction to regression (Lecture12) rather than the two scheduled. These adjustments to the course schedule are indicative of the professor's commitment to deep understanding of key statistical concepts over breadth of techniques initially determined to be valuable.

Foster active learning in the classroom. In the first interview Professor C's teaching style is self-described as being interactive, particularly through team exercises embedded in lectures. Ten of the 24 lectures include a slide that directs students to get into teams. The teams' work is either included in the lecture slides when the task is small (e.g., Lecture12) or a handout is given for the larger tasks or when raw data is needed (e.g., Lecture3). Lecture materials and observations confirm the professor's description and provide evidence that the course matches the GAISE suggestion to "mix lectures with activities, discussions and labs" (ASA, 2005, p. 18).

There are 11 lab assignments that students may complete collaboratively with classmates. A new feature of the lecture portion of the course this semester is the introduction of an iClicker system that brings further opportunities for the dual purpose of interactivity and formative assessment. The conversations illustrating Case C's stress on conceptual understanding in the previous section also indicate student willingness to engage in group discussion of interesting topics.

Another GAISE suggestion that Case C demonstrated is the collection of data from students. However, a further suggestion is to collect data in a context, with a question that the data can answer. The data collection from the student survey occurred in an unobserved lab session that may have had some verbal context given, but there is no written evidence among the course documents to suggest this. Professor C elaborated during the review of this analysis that the students suggested the variables and were asked for hypotheses, but they have a difficult time doing so that early in the semester (Personal communication, post-analysis). When the data is used in later lectures, a context is provided (e.g., the team activity in Lecture 21).

Use technology for developing conceptual understanding and analyzing data. The syllabus describes one of the course objectives as "learn how to manage data and conduct analyses using SPSS" (p. 1), the statistical software package frequently used in the discipline and in the professor's own research. GAISE recommends the use of such software for the purpose of allowing the course focus to be on the interpretation of results instead of computation, which is the exact purpose of most lab/SPSS assignments (e.g., Lab_11). Some students use a graphing calculator to conduct statistical analyses (Observation 6), though Professor C does not provide any instructions on how to use it.

GAISE also recommends, "Regardless of the tools used, it is important to view the use of technology not just as a way to compute numbers but as a way to explore conceptual ideas and enhance student learning as well" (ASA, 2005, 12). The lectures on descriptive statistics include histograms and column charts that were software-generated (e.g., Lecture5) as does the introduction to hypothesis testing with normal curves and their shaded tails (Lecture19). There is also a spreadsheet to demonstrate the variability in repeated samples as a precursor to sampling distributions (Observation 5). These do help in the visualization of concepts as GAISE suggests but fall short of being a way to "develop an understanding of abstract ideas by simulations" (p. 12) or "explore 'what happens if…'-type questions" (p. 13).

Use assessments to improve and evaluate student learning. Professor C includes a range of assessments throughout the semester: four exams, a cumulative final exam, homework, lab assignments, weekly quizzes, and a research proposal (Syllabus). On two occasions clicker quizzes allowed for informal assessment unrelated to the course grade (Lecture3, Lecture6). This variety matches the GAISE suggestion for a more thorough evaluation of learning.

The introduction of online homework allowed for "immediate feedback, the opportunity for multiple attempts—three times before the deadline—and you can re-do the problems even after the deadline but without any grade" (Observation 1).

> When I was controlling the homework myself, I didn't post the homework until I thought they were ready to complete it. With this I have to set the schedule at the beginning of the semester. I had to keep track that what was online aligned with what we were doing. It turned out to be more assignments, although shorter and more frequent. (Instructor interview, post-semester)

These features of the homework assessments make them a good fit for the GAISE suggestion that tasks should be well coordinated with topics in recent classes. Already mentioned in an earlier section is the inclusion of conceptual questions that the Professor admits to neglecting when the homework was self-designed (Instructor interview, post-semester).

Summary. Case C presents evidence for matching the Goals for Students in most areas listed by GAISE. The block of goals related to conducting procedures showed the largest number of matches. Goals for student understanding of the basic ideas of statistical inference also had many matches but the lack of instruction about confidence intervals weakens Case C's alignment with GAISE in this area. Critical thinking about statistical results and how they are reported, particularly in professional journals, is evident in Case C.

Regarding the GAISE Recommendations for Teaching, Case C shows convincing evidence of emphasizing statistical literacy and developing statistical thinking as well as stressing conceptual understanding, not merely knowledge of procedures. This case demonstrates active learning through in-class activities and the lab component of the course. The labs provide most of the evidence of the use of technology for developing concepts and analyzing data. The immediate feedback from online homework shows Professor C's interest in using assessments to both improve and evaluate student learning.

Case D – Statistics for Students in a Social Science Major

The setting. This course is a requirement for students majoring in a social science discipline different from the one in Case C. There are no mathematical pre-requisites for taking this course but a research methods course does precede this one.

116

Students generally take both courses in their junior year (Instructor interview, pre-semester). During the first class, as part of some informal data gathering, the professor remarked on the number of students responding to the request for a show of hands if they are in their third year of the program, "I'm glad to see seniors are not waiting to the last minute" (Observation 1).

There are 68 undergraduate students enrolled in the same lecture section, meeting twice a week for 50 minutes. Three graduate students are also enrolled in preparation for a course in multivariate regression required for their degree. Each student is also enrolled in one of the four two-hour lab meetings at the end of the week. Observations took place only in the lecture sessions. The classroom is furnished with fixed desks and moveable chairs in tiered ranks on either side of a center aisle. There is a computer podium for the instructor, three whiteboards across the front, which are partially covered by the screen when the projector is in use. There are approximately the same number of women and men enrolled in the course.

Professor D has taught this course "about four times in the past five years" (Instructor interview, pre-semester), the entirety of the professor's affiliation with the institution (Instructor CV). "That's why I got hired…to teach stats and methods. I was hired as a quant[itative] person…The first class I taught [here] was stats." (Instructor interview, pre-semester). "The department here is not very stats focused" (Instructor interview, post-semester).

Course design. There is an online course management system available to support the instructor and participants, including the observer. The presentation slides for the lectures (see examples in Appendix C), homework assignments, data files used for

homework or during lab sessions, and the syllabus are found in the resources section of the website. The system has functions for sharing grades, presenting and collecting assessments, and a discussion board but these are not utilized in this course.

Lectures did not strictly follow the textbook's order of presentation; completely skipping some chapters but including the optional chapter on Analysis of Variance (DeVeaux, Velleman, & Bock, 2006; Syllabus). The syllabus indicates the predetermination to skip the three chapters on data gathering and the one with non-normal probability models. At the end of the semester, Professor D explained that "the research methods class has a lot of the big ideas" about data and data collection (Instructor interview, post-semester).

There are two teaching assistants who each supervise two lab sessions each week. They attend the lectures as well as grade the homework, quizzes, and exams from the undergraduates. Both faculty and graduate students in the department make lists of preferences for graduate teaching assignments. "I get more input on who TAs for this class. It requires a particular skill set and level of commitment … I look for the ones who *want* to TA for this class and consider how they did in the grad stats course" (Instructor interview, pre-semester).

Assessments. This course assesses student learning through nine weekly homework assignments, three unannounced quizzes, two in-class exams, and a final take-home assignment that requires the use of SPSS. Two-thirds of the quiz or exam problems and homework exercises include multiple parts that assess different types of student understanding. Twenty-three percent of the questions, particularly in the earliest homework assignments, ask strictly procedural or identification questions such as "Name

118

the variables and explain whether they are quantitative (continuous) or categorical. If a

variable is quantitative, note its units; if it is categorical note whether it is nominal or

ordinal" (Homework 1). Five percent of the questions are strictly conceptual; for

example, "Assume that question #5 [a one-tailed test] asked: Is there evidence that

different proportions of women and men buy books on-line? Would your conclusion be

different? Why or why not?" (Homework 4).

The final take-home assignment is a collection of eight multi-part questions about

a data set the students have not used previously in the semester. "A key part of this take-

home is recognizing the right procedure to answer particular questions" (Observation 8).

An example of blending conceptual and procedural questions follows:

> Does students' academic self-confidence increase from 8th to 12th grades?
> a)   State appropriate hypotheses. (2 points)
> b)   List and check all appropriate assumptions. (4 points)
> c)   Conduct the appropriate test and copy the appropriate table from SPSS.
>   Using $\alpha=0.05$, state your conclusion statistically and in context. Make
>   sure to be explicit about p and alpha values you are using to make your
>   conclusion. (6 points)

Other research questions ask students to "report and interpret the appropriate confidence

interval," either in addition to or replacing the hypothesis test. Question 4 adds some

complexity to the task with a follow-up question to the hypothesis test: "If you were a

policy maker who wanted to improve math scores, what track would you recommend

students enroll in and why?" Quiz and exam questions are much like the take-home

assignment but require hand calculation for single sample tests or provide the SPSS

output to interpret.

Case D – Statistics for Students in Social Science Majors, Pattern Matching Analysis

Professor D did not know about GAISE "but I'm not surprised" (Instructor interview, post-semester). As the following analysis indicates, the instructor's lack of awareness does not prevent Case D from demonstrating some of the same goals for students or implementing the recommended pedagogy.

Goals for students. During lectures, the instructor's verbal remarks concerning topics listed in the five blocks of goals for students were tallied on the observation protocol (see Appendix A). Coding of interview transcripts and observation notes supplement the tallies on the protocol to provide frequency counts of Verbal evidence that are reported in the tables accompanying the analysis. The frequencies of Written evidence come from the coding of lecture notes and the syllabus. Since the textbook was not available electronically for NVivo coding, the frequencies do not include this major source of written evidence. Coding of exams, quizzes, homework, and the final take-home assignment provides the counts in Assessed column of the tables.

The initial look at how Case D's goals for students align with the five blocks of the GAISE list (see Tables 19 through 22) shows minimal evidence in the first two blocks, plentiful evidence in the third and fourth with some evidence in the fifth.

First block of goals. The lack of counts in table 19 suggests that Case D is not concerned with student understanding of concepts about what information statistical analysis can and cannot provide. The one item counted comes from the pre-semester interview where Professor D says, "They should know that statistical significance does not necessarily make something meaningful. Even if you have a big enough sample, and everything is significant, is it really meaningful?" However, without corroborating

evidence that this was ever communicated to students, it does not provide confidence that these goals apply to Case D.

Table 19 First Block of Goals, Coding Frequencies			
Students should believe and understand why…	Verbal	Written	Assessed
Data beat anecdotes	0	0	0
Variability is natural, predictable, and quantifiable	0	0	0
Random sampling allows results of surveys and experiments to be extended to the population from which the sample was taken	0	0	0
Random assignment in comparative experiments allows cause-and-effect conclusions to be drawn	0	0	0
Association is not causation	0	0	0
Statistical significance does not necessarily imply practical importance, especially for studies with large sample sizes	1	0	0
Finding no statistically significant difference or relationship does not necessarily mean there is no difference or no relationship in the population, especially for studies with small sample sizes	0	0	0
Note: Cumulative frequency does not match NVivo count because some items matched multiple goals. Also, the Written column is undercounted because the textbook was not available electronically.			

Two important factors that are not reflected in Table 19 are the textbook and the research methods course that precedes this one. Professor D is quoted above describing the prerequisite course as being focused on processing data. That course description mentions "conceptualization of social problems" and "emphasis on student projects" as well as "data processing." The intentional neglect of the textbook chapters on data gathering supports the professor's supposition that these goals have already been addressed in the previous semester.

121

It should also be noted that, even without the chapters on data gathering, the textbook addresses most of the GAISE goals in this block at least once. This is not surprising since the preface to the text includes the following:

> We have worked to provide materials to help each class, in its own way, follow the guidelines of the GAISE (Guidelines for Assessment and Instruction in Statistics Education) project sponsored by the American Statistical Association. (DeVeaux, et al., 2006, p. xiii)

Every chapter of the textbook has a section titled "What Can Go Wrong?" Looking only at this section and only for the chapters listed in the syllabus, all but "variability is natural, predictable, and quantifiable" receives further explanation (DeVeaux, et al., 2006). Variability is a key theme in the introductory pages, summarized with the statement, "Statistics is about variation" (DeVeaux, et al., 2006, p. 3). The chapter on correlation includes a lengthier discussion of association and causation; significance versus importance has its own section in the chapter on inference about means (DeVeaux, et al., 2006). Student exposure to these ideas, however, depends entirely on their diligence in reading the text; in the post-semester interview Professor D expresses the suspicion that they do not read the text.

Second block of goals. Like the previous block, Professor D indicates that these are part of the research methods course. Table 20 shows a similar dearth of evidence of these goals for students, but the textbook provides less uncounted support than it did for the previous block.

Table 20 Second Block of Goals, Coding Frequencies			
Students should recognize…	Verbal	Written	Assessed
Common sources of bias in surveys and experiments	0	0	0
How to determine the population to which the results of statistical inference can be extended, if any, based on how the data were collected	1	0	0
How to determine when a cause-and-effect inference can be drawn from an association based on how the data were collected (e.g., the design of the study)	0	0	0
That words such as "normal," "random," and "correlation" have specific meanings in statistics that may differ from common usage	2	0	0
Note: Cumulative frequency does not match NVivo count because some items matched multiple goals. Also, the Written column is undercounted because the textbook was not available electronically.			

The three chapters of the textbook that are not included in Case D's syllabus do address the first three goals in this block. The mathematical/statistical use of the word of "random" is also included at the beginning of that part of the text (DeVeaux, et al., 2006, p. 251). A curious student would have a resource for learning about these important ideas even though they are not included in the course.

The two instances counted regarding the use of words are both statements by Professor D regarding the meaning of "significant." The first mention is at the introduction to hypothesis test: "Significant -- statistically this means something very unique. It's not what we mean in lay language" (Observation 3). A similar statement comes in the next lecture: "People use the word significant all the time without any particular precision. When we say statistically significant we precisely mean that the P value is less than α" (Observation 4).

123

Both "normal" and "correlation" receive attention by the textbook authors to warn students that their precise meaning in statistical context differs from their everyday meanings. "'Normal' doesn't mean that these are the *usual* shapes" (DeVeaux, 2006, p. 106) and "Don't say 'correlation' when you mean 'association'" (p. 152). In each case further information about the distinctions is provided. None of this careful vocabulary is assessed.

Third block of goals. Table 21 summarizes the evidence that Case D shares GAISE goals regarding procedures for obtaining and analyzing data with appropriate techniques and meaningful communication of the results.

Table 21 Third Block of Goals, Coding Frequencies			
Students should understand the parts of the process through which statistics works to answer questions…	Verbal	Written	Assessed
How to obtain or generate data	0	0	0
How to graph the data as a first step in analyzing data, and how to know when that's enough to answer the question of interest	1	5	4
How to interpret numerical summaries and graphical displays of data—both to answer questions and to check conditions (to use statistical procedures correctly)	2	5	2
How to make appropriate use of statistical inference	1	6	5
How to communicate the results of a statistical analysis	3	3	6
Note: Cumulative frequency does not match NVivo count because some items matched multiple goals. Also, the Written column is undercounted because the textbook was not available electronically.			

The first item in this block of goals for students lacks evidence outside the chapters/topics not covered in Case D, though the remaining goals are evident. The second lecture of the semester is the only class time devoted to the *how* part of graphing or calculating summaries of data. The subsequent lecture moves on to the interpretation

124

of graphs and numerical measures for answering questions. Lecture 3 also uses these summaries to check conditions (i.e., normality) for the first time. "Since all of our procedures will depend on the normal model, we always have to check that the data is normally distributed before using techniques" (Observation 2). A later class reminds students to "check histogram or P-P plot" before completing a *t*-test of the mean

> Look at descriptive statistics for an initial look at the data … just because there is a mathematical difference doesn't mean they are statistically different. Notice the high standard deviation in the male group. It's important to look at the descriptives first to get a sense of what the data actually looks like. (Observation 6).

Student understanding of using graphs to answer questions is assessed by comparative box plots to answer questions (Homework 1) and discuss symmetry of the data (Quiz 1). Scatterplots are used to ask if correlation (Homework 8) or linear regression (Homework 9) is an appropriate analysis. The final take-home assignment asks for assumptions to be both listed and checked, "include the histogram or the table to show that the data fits the assumptions" (Observation 8).

Most of the evidence for the goal of making appropriate use of statistical inference was also coded in the previous goal because of checking conditions. Two instances that were more complex follow:

> With more than 2 groups, why not just run multiple t-tests?
> - Probability of making Type I error will exceed the chosen α
> - Family-wise error related to the complete set of comparisons will be $k * \alpha$ (simple formula)
> - If you wanted overall $\alpha = 0.05$, each test would need to be based on α/k (k=number of comparisons)
>
> To keep Type I error at a specific α-level regardless of the number of comparisons – ANOVA
> (Lecture 14, slide 3)

The in-class activity during Observation 8 provided scenarios and research questions for which students name the appropriate test and why. For example,

> Researchers want to measure the effect of divorce on educational success in high school. They randomly select 2000 high school students and divide the sample into two groups: those with divorced parents and those with married parents. They then record the GPA of each student. Do children of divorce have lower GPAs than children of married parents? (Linking Research Questions 2 Tests).

Professor D models good communication of statistical results, always stating conclusions in context and explicitly describing the expectations for how to do this. "The statistical conclusion is always about null. The contextual conclusion is always about the alternate [hypothesis]" (Observation 4). The introduction of both hypothesis testing and confidence intervals uses the same example of a question about whether binge drinking at "your" school is higher than the national average. The model conclusions are below.

> Reject the null. There is evidence that binge drinking at your school is higher than the national average (Lecture 7).
> Reject the null. We are 90% confident that between 45% and 55% of college students engage in binge drinking (Lecture 8).

The same models apply to two sample tests and extended to other tests.

Careful interpretation of correlation and regression analysis is also modeled and emphasized in Lectures 16, 17, and 18. Stating conclusions and interpreting statistical results are repeatedly assessed through homework exercises (e.g. Homework 5), quiz and exam questions (e.g., Quiz 2 and Exam 2), and in the final take-home assignment.

Case D shows indicators that students learn to use descriptive statistics, including graphs and exploratory data analysis, for answering questions and checking conditions. There is emphasis on the selection of appropriate inferential procedures and effective communication of results.

Fourth block of goals. Important concepts related to the need for accurate

interpretation of inferential analysis are the goals in this block (see Table 22).

Table 22 Fourth Block of Goals, Coding Frequencies			
Students should understand the basic ideas of statistical inference…	Verbal	Written	Assessed
The concept of a sampling distribution and how it applies to making statistical inferences based on samples of data (including the idea of standard error)	2	2	1
The concept of statistical significance, including significance levels and p-values	3	3	8
The concept of confidence interval, including the interpretation of confidence level and margin of error	1	3	10
Note: Cumulative frequency does not match NVivo count because some items matched multiple goals. Also, the Written column is undercounted because the textbook was not available electronically.			

Lecture 6 is devoted to the concept of sampling distributions with specific

reminders at the later introduction of the *z*-test of a proportion (Lecture 7; Observation 3)

and confidence intervals (Lecture 8; Observation 4). Professor D describes the sampling

distribution as a "bridge between the sample information and the population. Remember

what a sampling distribution is: repeated sampling, plotting all possible sample means"

(Lecture 7; Observation 3).   The textbook devotes a chapter to the topic, asking readers

to "imagine the results from all the random samples of size 1000 that we didn't take" or,

better yet, "simulate a bunch of those random samples of 1000 that we didn't really

draw" (Deveaux, et al., 2006, p. 406-7).   Standard error is defined and calculated in

lectures and the textbook without derivation. No students ask for more explanation but

they are able to supply the necessary values during example calculations during

subsequent lectures. "You should be dreaming this formula by now," the Professor

remarks as the class constructs a confidence interval (Observation 5). Exam 2 includes a

question asking for a description of the sampling distribution to be referenced in making inference about a proportion.

Significance levels or "alpha levels" (Lecture 7; Observation 3; Observation 4) are described in lectures as the value against which the observed P-value is compared for making a statistical decision. Three commonly used values are discussed with examples of when each would be appropriate. "These values are the probability of error you are willing to tolerate" (Observation 4). Further discussion of errors and α as the probability of rejecting a null hypothesis that is true takes place in Lecture 12. There are no direct assessment questions regarding significance but there are occasions where students are asked about the effect of changing the value of α on confidence interval: "Would a 90% confidence interval have a smaller margin of error?" (Homework 4).

P-value receives a careful definition when introduced during the first lecture on hypothesis testing:

> Be careful about interpreting the p-value. If p-value is 2.5%:
> - It does NOT mean that H_0 is true 2.5% of the time
> - It does NOT mean that you are 2.5% certain that H_0 is true
> - It means that, given the null hypothesis, there is a 2.5% chance of observing the statistic value we actually observed (or higher).
>
> (Lecture 7)

During class the professor clarifies the "or higher" remark on the slide as being the case since the example is a right-tailed test but that in general it refers to "being more extreme; farther from the mean" (Observation 3). The textbook emphasizes the need for reporting the precise p-value "to show the strength of the evidence against the hypothesis. This will let each reader decide whether or not to reject the null hypothesis" (DeVeaux, et al., 2006, p. 459). Case D assessments ask, "Make sure to be <u>explicit about p and alpha</u>

<u>values</u> you are using to make your conclusion" (Take-home assignment; emphasis in the original).

The modeling of communication of statistical results discussed as part of the third block of goals with the example about binge drinking also illustrates the emphasis on interpretation of a confidence interval. The homework question quoted above as evidence of understanding the connection between significance and confidence also serves to support Case D's goal for students to understand margin of error. Homework 5 has the only explicit assessment of a margin of error: "What is the margin of error for this confidence interval?" when the interval (29.202, 31.844) is given. Homework 6 adds a question assessing student ability to apply the information learned through the construction and interpretation of a confidence interval: "What advice would you give to the company about framing its ad?"

Case D contains a range of evidence for its goals regarding student understanding of sampling distributions, statistical significance, and confidence intervals as the basic ideas of statistical inference. The course is in accord with the fourth block of goals for students outlined in GAISE.

Fifth block of goals. The last block of goals (see Table 23) relates to critical thinking about statistical results.

Table 23 Fifth Block of Goals, Coding Frequencies			
Finally, students should know…	Verbal	Written	Assessed
How to interpret statistical results in context	3	7	6
How to critique news stories and journal articles that include statistical information, including identifying what's missing in the presentation and the flaws in the studies or methods used to generate the information	3	0	0
When to call for help from a statistician	0	0	0
Note: Cumulative frequency does not match NVivo count because some items matched multiple goals. Also, the Written column is undercounted because the textbook was not available electronically.			

Professor D is consistent in providing or asking students to provide an interpretation of all inferential results. The example of a *z*-test of a proportion—the first inferential procedure of the course—sets the model for all subsequent procedures by making a contextualized interpretation an expected part of completing the hypothesis test. The last bullet on the lecture slide for stating a conclusion says, "state the conclusion in context: There is evidence that binge drinking at your school is higher than the national average (i.e., that the reputation as the 'party school' is justified)" (Lecture 7). The textbook also links the statistical conclusion with the contextual conclusion, "as always, the conclusion should be stated in context" (DeVeaux, 2006, p. 454). All of the assessments regarding inferential procedures ask students, "What is your conclusion, stated statistically and in context?" (e.g., Exam 1).

In the first interview Professor D expressed the expectation that students should complete the course "able to pick up the paper or report with basic descriptives or statistical claims and they should know what questions to ask. A critical eye should be automatic." Asked at the end of the semester if this goal was reached by the students:

I think lots of them do. I sometimes get e-mails from students sharing published claims and their critiques. I want them to develop that skepticism. They may not know exactly what is wrong but they are asking questions.

On the first day of class, Professor D tells the students,

I would argue that to be an educated citizen in the 21st century in America, you need to have a basic understanding of statistics…. You need to understand how and why people are making certain claims... They look so appealing, objective; they are numbers, right? They look so real. When they get misused, some people do it purposely but most of the time it is because they don't understand statistics.

The textbook supports the Professor's argument succinctly: "Always be skeptical" (DeVeaux, et al., 2006, p. 14). There is no assessment of whether the students have learned this skepticism other than the delayed response reported by the professor in the quote above.

Calling for help from a statistician is implied when the professor comments, "In real life, if you don't meet the assumptions, there are other tests you can use" (Observation 5).

There is some evidence that Case D shares the GAISE goals for students listed in this block. Alignment is strongest for interpreting results in context with multiple verbal remarks about critiquing published statistics. There is no evidence concerning the need for expert help.

Recommendations for teaching. Review of the observation protocols (see Appendix A) provided the frequency counts listed in the Verbal column of Table 24. Frequency counts from NVivo informed the Written column (lecture notes and syllabus) and Assessed (homework, quizzes, exams, and final take-home project).

Table 24 Recommendations for Teaching, Coding Frequencies			
	Verbal	Written	Assessed
Emphasize statistical literacy and develop statistical thinking	7	9	5
Use real data	0	1	0
Stress conceptual understanding, not merely knowledge of procedures	7	2	9
Foster active learning in the classroom	7	3	0
Use technology for developing concepts and analyzing data	0	13	8
Use assessments to improve and evaluate student learning	0	0	0
Note: Frequency of Written occurances do not include the textbook.			

Emphasize statistical literacy and develop statistical thinking. The syllabus for

Case D sets up the tone for the course:

> In this course, you will learn how to use statistics to understand everyday events, examine patterns in social life, evaluate claims, and develop a healthy skepticism for conventional wisdom and popular opinion. As such, this course focuses on developing analytical skills and learning to see the world through a statistical lens.

Professor D tells students on the first day that "statistics is a language that uses numbers

to talk about the world. A way to understand the world, a way to interpret the world. … A

tool for thinking about the world" (Observation 1). These statements are evidence that

statistical literacy and thinking are important objectives of the course aligning with the

GAISE definition of statistical literacy: "understanding the basic language of statistics …

and understanding some fundamental ideas of statistics (ASA, 2005, p. 14).

As mentioned earlier, the authors wrote the textbook with the GAISE College

Report in mind. This orientation to teaching statistics is evident in the way that the

mathematics is handled. "The equations we use have been selected for their focus on

understanding concepts and methods" (DeVeaux, et al., 2006, p. xiv). Formulas are almost non-existent in the lecture notes. When they are necessary, they are written in words, with a minimum of mathematical symbols. For example, the "formula" for a confidence interval is given as "Estimate ± margin of error" followed by the margin of error defined as "z * SE…SE formula on pg. 495" (Lecture 9, slide 8). Standard error formulas for single sample tests are the only ones written mathematically and only on the whiteboard (Observation 3; Observation 4). More complex formulas are neglected entirely, substituted by output tables from SPSS (e.g., Lecture 11 and Lecture 15).

The final assignment in the semester requires the statistical thinking necessary to choose an appropriate test procedure for answering the research questions posed. Throughout the semester, Professor D models statistical thinking, as GAISE suggests, through the lecture examples and their accompanying explanations. However, the prior homework, quiz, and exam assessments have directed students to the procedure by either explicitly specifying it or by providing SPSS output for interpretation, which is opposite to the GAISE suggestion. Some consideration in the earlier assessments for students' ability to choose the correct procedure would enhance Case D's alignment with this recommendation.

Use real data. "The students use real data in the labs with the TAs. We work with the summary statistics in lecture" (Instructor interview, post-semester). Since lab sessions were not observed, the evidence for student use of real data sets in the course is weak but does exist through the data sets used for assignments. The course management system includes four data sets in SPSS format. Two of these, with 160 and 400 observations, are required for homework sets 5 through 9; the largest set, with 500

observations, is required for the final take-home assignment; and, the smallest set of just 26 observations is not mentioned in any Case D documents, possibly for use in labs. The two exams are prefaced with the statement "Note that the examples are developed for illustrative purposes only and may not reflect actual data or relationships."

There is stronger evidence that Case D agrees with the GAISE suggestion to "make sure questions used with data sets are of interest to students" (ASA, 2005, p. 16). Many of the inferential procedures are introduced in lecture using a question about the characteristics of college students (e.g., Lecture 10 and Lecture 15). Of particular interest to students at the end of their third year in college is the prediction equation for wages based years of education (Lecture 17).

Stress conceptual understanding rather than mere knowledge of procedures. In the first interview Professor D expressed commitment to conceptual understanding for the students:

> I stress the logic over the math – statistical reasoning. …. I'm much less strict or concerned about coverage, more about the conceptual/logic. … I like the second half of the semester a lot more. It's because they can think and do stuff on their own. Like the independent and paired *t*-test, once they've done a one sample *t*-test. I can almost let them do it for themselves, even the first time. They can start figuring out on their own. They can be more engaged then because they have enough background.

Students are told the same thing in the first class: "This class does have math – we can't get away from the math entirely – but it is not a math course. We will focus on the logic" (Observation 1). Describing the course structure is another opportunity for Professor D to reiterate the theme by saying, "Classes will be mostly lecture about the concepts. … Labs will be run by the TAs and focus on the use of SPSS as well as homework help. Conceptual ideas in lectures, applications in lab" (Observation 1).

Keeping formulas in words rather than symbols is one way that Professor D focuses attention on the concept of hypothesis testing. "Think logically. How are we going to calculate t? Sample minus population divided by standard error—always the underlying principle. How does this translate to two samples?" (Observation 6). Questioning the students in this fashion increases the student engagement mentioned in the pre-semester interview and addresses "a key part in teaching intro stats is to get students to figure out that a) it's not terrible and b) it's not terrifying… I hope they start to see that they know the logic and can muddle through additional tests" (Instructor interview, pre-semester).

Using SPSS for the computations is also part of Professor D's strategy to keep the focus on the concepts. This fits nicely with the GAISE suggestion for "using technology to allow greater emphasis on interpretation of results" (ASA, 2005, p. 18). SPSS output is presented on lecture slides eliminating in-class computations once the basic inferential concepts are covered with single sample z and t tests. "They complain so much about the math. … Where is all this math?" (Instructor interview, post-semester).

All assessment avenues—homework, quizzes, exams, and the final take-home assignment—include directives to "explain your answer" (e.g., Exam 1) or "be explicit about values and logic used to make your decision" (e.g., Quiz 2) and ask questions like "What type of a test is needed to test your hypotheses? Explain" (e.g., Homework 6). A few homework questions assess particular concepts, such as "Why doesn't the model explain 100% of the variation in the price of an Escort?" (Homework 9) and "What happens to the correlation if income is measured in thousands?" (Homework 8).

Case D shows evidence that Professor D's explicit intention of focusing on understanding concepts carries through lectures and assessments.

Foster active learning in the classroom. Professor D describes the teaching style as "engaged lecture" (Instructor interview, pre-semester). As predicted by the instructor, students are more engaged once they pass t-tests because Professor D elicits their knowledge about the structure of hypothesis testing to reason through additional procedures (Observation 5; Observation 6; Observation 7). While not quite the "problem solving, activities and discussion" advocated by GAISE, it does demonstrate that the professor does not "overestimate the value of lectures" (ASA, 2005, p. 18).

"The lab is where the students work in teams to solve problems and share their solutions with the larger group. We want that to be problem-solving based" (Instructor interview, post-semester). Since the labs were unobserved, there is no direct evidence to support Case D's interest in active learning through the lab sessions. The instructor-reported pass rate for the course (90%) may be taken as proxy evidence that the students attended and participated in lab—10% of the course grade—in which they learned to use SPSS for answering the statistical questions on the final take-home assignment worth 25% of the overall course grade (Syllabus).

On the first day of the semester, Professor D took an informal poll of the students regarding their major, their home state, and their year in college (Observation 1). Results were not recorded, which prevents this activity from constituting evidence of the GAISE suggestion to "collect data from students" (ASA, 2005, p. 19). It is a missed opportunity for Case D "to take advantage of the fact that large classes provide opportunities for large sample sizes for student-generated data" (ASA, 2005, p. 19).

Further investigation of the lab sessions may provide the evidence that is lacking for Case D's use of active learning.

Use technology for developing conceptual understanding and analyzing data. When asked about the use of technology in the course, Professor D mentioned the use of output from SPSS instead of tedious hand calculations (Instructor interview, pre-semester). All inferential procedures involving two or more populations or bivariate data are presented with SPSS output instead of providing formulas for hand calculations (Lecture 11 and following; Observation 6; Observation 7). This is in perfect alignment with the GAISE recommendation to use software for computation in order to allow students to focus on the interpretation of results.

GAISE also recommends, "Regardless of the tools used, it is important to view the use of technology not just as a way to compute numbers but as a way to explore conceptual ideas and enhance student learning as well" (ASA, 2005, 12). Scatterplots (e.g., Lecture 18), histograms (e.g., Lecture 13), and box plots (e.g., Lecture 14) needed for checking conditions are software-generated but static in the lecture notes. These do help in the visualization of concepts as GAISE suggests but falls short of being a way to "develop an understanding of abstract ideas by simulations" (p. 12) or "explore 'what happens if …'-type questions" (p. 13).

Use assessments to improve and evaluate student learning. The nine homework assignments "are well coordinated with what the teacher is doing in class" (ASA, 2005, p. 21), therefore, they are expected to be effective learning tools. The three unannounced quizzes are also aligned with the course's current topics. Two exams and the final take-home project—described by Professor D as "like an SPSS exam" (Observation 1)—

complete the variety of assessments used in the course. The tasks in each type of assessment are not especially different in terms of cognitive complexity, so it could be argued that they do not meet the GAISE suggestion for "a variety of assessment *methods* to provide a more complete evaluation of learning" (ASA, 2005, p. 21, emphasis added). However, the mixture of procedural and conceptual knowledge required to answer the majority (67%) of questions across all types of assessment does assess "understanding [of] key ideas and not just on skills, procedures, and computed answers" (ASA, 2005, p. 21).

The TAs grade all the assessments and the quality of their feedback was not observed in this study. Further investigation is needed to know how feedback may play a role in Case D. There are no assessment items asking for interpretation or critique of the use of statistics in popular media with which to evaluate statistical literacy goals, nor are there projects or investigations to assess statistical thinking, leaving Case D with little evidence for "assessments [that] lead to learning" (p. 13).

Summary. Case D shows little evidence for matching the Goals for Students in first two blocks listed by GAISE. The blocks of goals related to conducting procedures and student understanding of the basic ideas of statistical inference had far more evidence of Case D's alignment with GAISE. Interpreting statistical results in context provided most of the evidence for the final block of goals.

Emphasis on statistical literacy and development of statistical thinking as well as focus on conceptual understanding, not merely knowledge of procedures are areas where Case D aligns well with the GAISE Recommendations for Teaching. The use of SPSS—output in lectures, student use for assessments—provides most of the evidence for the use

of technology for developing concepts and analyzing data. Active learning is the

intention for lab sessions, with efforts at dialogue with students during lecture providing

the observable evidence. Using real data and using assessments to improve learning are

not evident in Case D.

Chapter 5: Cross-Case Analysis

The previous chapter addressed each case individually, exploring the question of

how GAISE goals and recommendations are evident in a variety of settings. This chapter

will consider areas where the cases align with GAISE similarly and where they differ in

their alignment. The analysis will begin with a comparison of the administrative

structures of the cases and some brief comments on the variety represented by these four

cases. There will be separate analyses of the cases' goals for students and the pedagogy

used by each. The chapter will conclude with a discussion of the themes found within

and across the cases.

Variety of Course Structures

When George Cobb led a focus group of tertiary educators interested in the

introductory statistics course and subsequently published a report to the Mathematical

Association of America (MAA) in 1992, much of the diversity that the GAISE College

Report calls "a family of courses" (ASA, 2005, p. 7) was clearly evident. The MAA

focus group "made a deliberate decision not to prescribe lists of topics … instead to seek

a general intellectual framework within which we and others can fit a great variety of

courses" (Cobb, 1992, p. 1). Some of the structural variety reported in 1992 and

reiterated or revised in 2005 is described as

> Calculus prerequisite versus no calculus; engineering, technical audience versus
> arts, nontechnical audience; goal of understanding versus goal of doing; taught by
> mathematics or statistics department versus taught by user department; large
> research university versus small college; large clientele (100s – 1000s) versus
> small clientele (less than 100); required course versus elective course; students

bright, intellectually curious versus students dull, passive; PC's readily available versus computer facilities inadequate. (Cobb, 1992, p. 1)

The GAISE College Report added the possibility of distance learning settings and considered the length of course (weeks in a semester and time in class each week).

The four cases analyzed individually in the previous chapter reflect the many variations in the administrative structure that Cobb and GAISE acknowledge (see Table 25). All four cases take place in a 15-week semester, in face-to-face classrooms, and the students are required to take this course as part of their major; they are otherwise quite diverse.

Table 25 *Matrix of Case Descriptions*	Case A	Case B	Case C	Case D
Class size	200[a]	453[b]	33	68
Pre-requisite	Calculus	Calculus	None	Research Methods
Majors	STEM	Business	Social Science	Social Science
Instructor background	Discipline	Stats	Discipline	Discipline
Instructor experience	~10 years	Twice	~6 years	4 times
Support personnel	3 TAs	8 TAs	1 Tutor	2 TAs
Lab for software use	N	N	Y	Y
Software	Minitab	Excel	SPSS	SPSS
Hours per week	3 - lecture	1 ¼ - lecture 1 ¼ - recitat.	3 - lecture 2 - lab	2 - lecture 2 - lab

Notes: [a] Three sections with approximately 70 student in each. [b] Three sections with approximately 150 students in each.

Half the cases require calculus, though Case B does not use it at all; the other half requires weekly lab sessions using SPSS. There is a range of class sizes, majors, contact hours per week, instructor experience, and support staff. In three of the four cases, the instructor's background matches the student major, the exception being the case with the largest number of students and the shortest time spent in contact with students.

Mathematics. GAISE is intentionally silent on the subject of calculus as a prerequisite for an introductory statistics course. The need for calculus depends on the topics covered, "we are not recommending specific topical coverage" (ASA, 2005, p. 11). Case A includes two such topics, while Case B does not.

The textbook in Case A leans heavily on calculus in the discussion of continuous random variables. Mean, median, percentile, and variance are redefined in terms of the area under a curve, using integration of functions that are decidedly non-normal. One such curve models the time between emission of alpha particles for a certain radioactive mass: $f(x) = 0.1e^{-0.1x}$ for $x > 0$ (Navidi, 2011, p. 106). The subsequent chapter on the propagation of error depends on evaluating derivatives and also includes an interesting perspective regarding the standard deviation of the sampling distribution of sample means:

> With a little thought, we can see how important these results are for applications. What these results say is that if we perform many independent measurements of the same quantity, then the average of these measurements has the same mean as each individual measurement, but the standard deviation is reduced by a factor equal to the square root of the sample size. In other words, the average of several repeated measurements has the same accuracy as, and is more precise than, any single measurement. (Navidi, 2011, p. 165-166)

These two topics treated through the calculus are of discipline-specific importance and the first exam assesses student ability to perform the calculations. "Measurement is

fundamental to scientific work. Scientists and engineers often perform calculations with

measured quantities" (Navidi, 2011, p. 157).

During the initial interview, Professor B articulated a different justification for the

calculus prerequisite:

> There isn't very much actual calculus, not the techniques of calculus. I'll talk
> about some of the concepts … about how integration is area under the curve but
> they are not required to evaluate an integral … I don't have the expectation that
> they can use the techniques of calculus but they should have a familiarity with
> math. They should have a certain level of confidence with math, thinking
> mathematically and doing mathematical problems.

Neither the textbook nor the professor in Case B demonstrates the calculus of probability,

but they both hint at it. "When necessary, we can once again call on more advanced

mathematics to learn the value of the standard deviation. The study of mathematical

methods for doing calculations with density curves is part of theoretical statistics … we

often make use of the results of mathematical study" (Moore, et al., 2009, p. 54). The

lecture notes for Topic 1 identify the standard deviation with the *inflection points* on the

normal curve, terminology not used by the textbook but familiar to students who have

studied calculus.

Cases C and D expect students to come with limited mathematical skills,

presenting a challenge not faced by the others. Professor C says, "A lot of them come in

with a phobia about math. I really do try to calm the phobia about math" (Case C

Instructor interview, pre-semester), and gives an assessment of basic arithmetic and

algebra skills in the first lab. Professor D also says, "They're [in this] major, in part,

because they didn't want to take math. Oftentimes they are appalled that they have to

take stats" (Case D Instructor interview, pre-semester). Both instructors assure their

students that the focus is on the concepts and that the tedious math will be done by the computer (Case C Observation 1; Case D Observation 1).

Software. SPSS (originally, Statistical Package for Social Science) is the obvious choice for use in Cases C and D considering the "ease of use for particular audiences" and "availability to students" (ASA, 2005, p. 21). Likewise, Excel is a natural choice for a course required for business majors. "I use Excel to handle some of the statistics functions. I know some people use calculators but I want something that everybody can use" (Case B Instructor interview, pre-semester). The textbook for Case A comes packaged with a student version of Minitab; neither the author nor the professor offers any justification for this choice.

All four cases use computer output during lectures on at least one occasion (e.g., ANOVA tables) and encourage—or require—their students to use software for computations in lab assignments (Case C and D), homework (Case B and D) or projects (Case A). Professor B is the only one to use software for analysis "live" in a lecture, though Professor C does demonstrate with SPSS during lab sessions. Students in Case A never see a demonstration but receive lots of written guidance; Case D students receive direct instruction from the TAs.

GAISE recommends that "technology tools should also be used to help students visualize concepts and develop an understanding of abstract ideas by simulations" (ASA, 2005, p. 19). Each case includes static representations of graphical summaries of data to illustrate abstract ideas (e.g., sampling distribution) in their textbooks and lecture slides. Only Case B provides a dynamic demonstration during class: an Excel spreadsheet that is also available to students for their own investigation outside of class (see Appendix D).

Instructor background and support personnel. It is no surprise that the case with the largest enrollment also had the largest contingent of supporting personnel and that the smallest class had only a single tutor assigned to the course. The connection that needs more investigation is that the largest class also had the fewest hours of contact with the professor and the professor did not share the students' discipline, while the smallest class had the most contact hours between students and professor in the same discipline. When asked if the goals for the course were achieved by most students, Professor C replied, "I think so … nobody is coming back [to repeat the course] next semester" (Case C Instructor interview, post-semester). Professor B did not provide any indication about the overall pass rate for the course to allow for comparison. This pair of strikingly dissimilar courses would make an interesting starting point for a study of student outcomes, both academic and attitudinal.

Case C also stands in contrast to the others by not having any help with grading students' written work. Professor A graded a quarter of the exams and projects, an even share with the TAs, but none of the homework or written quizzes. The online system did the homework grading and an optical mark recognition system graded exams for Case B. All undergraduate assessments in Case D were graded by TAs. Further discussion of these differences is part of the analysis of the recommendations concerning assessment later in the chapter.

Pattern-matching across cases. After completing the individual case analyses, each case received one-word descriptors for its alignment with GAISE's five blocks of goals for students and the six recommendations for teaching. Table 26 is a matrix of the goals by the four cases. Table 27 is a matrix of the Recommendations for Teaching by

the four cases. Together they provide a framework from which to consider how GAISE applied across the cases.

There is certainly some subjectivity involved in assigning these labels but the following definitions were in mind when applied:

Not evident – little or no attempt to include

Potential – little or no attempt but opportunity to do so

Uneven – some evidence for all parts or evidence for some parts

Aligned – evidence for most parts

Well-Aligned – multiple sources of evidence for all parts

The "not evident" label only applied to the first two blocks of goals in Case D where the professor explicitly said that these goals belonged to the research methods course. "Potential" applied where goals were evident from either the professor or textbook but not corroborated by the other (and not assessed) or where a small change to the course would initiate evidence of a teaching strategy, such as adding citations to the lecture notes when real data is used for examples. Designating a case/goal as "uneven" came from evidence for some but not all entries in a block or a mix of goals with corroboration but not triangulation; case/teaching designations of "uneven" resulted from inconsistent use during the semester, such as the use of think-pair-share activities in Case A. "Aligned" applied where evidence was triangulated on most goals or where a teaching strategy matched more than one of the suggestions in GAISE. When an excess of evidence existed, it was designated as "well-aligned."

The frequency counts in Tables 1 through 24 informed the initial labeling but evidence not available electronically for coding in NVivo—thus, not included in those

counts—prompted adjustments. For example, the fifth block of goals for Case C has zero

frequencies in four of the nine cells on Table 17 (page 106). However, uncoded evidence

from the textbook is presented in that case analysis that covers two of those zeros so that

the case/block gets labeled "aligned" rather than "uneven" as the descriptions in the

previous paragraph would designate.

Variety in Setting Goals for Students

Among the Goals for Students, all four cases showed alignment with GAISE in

the third and fourth blocks, both of which are related to statistical procedures. The other

three blocks have greater variety of alignment across the cases.

Table 26 *Matrix of Goals for Students*				
	Case A	Case B	Case C	Case D
First Block: concepts about what information statistical analysis can and cannot provide	uneven	well-aligned	uneven	not evident
Second Block: recognition of appropriate interpretation of results from statistical analysis	potential	aligned	uneven	not evident
Third Block: parts of the process through which statistics works to answer questions	aligned	well-aligned	aligned	aligned
Fourth Block: basic ideas of statistical inference	well-aligned	well-aligned	aligned	well-aligned
Fifth: critical thinking about statistical results	potential	potential	aligned	uneven

First and second blocks In the individual analysis, it became evident that within

Case D no effort was devoted to these goals. The instructor is confident that the ideas

that GAISE presents in these block are covered in the research methods course that is

prerequisite. Since there is no evidence to support or refute that claim, Case D is not

included in the analysis of these two areas.

The first block lists goals for students' beliefs and understanding of concepts about what information statistical analysis can and cannot provide. The one goal that received roughly equal attention across the three cases is "association is not causation," while the others were most evident in Case B with varying levels of agreement with one or the other of the remaining two cases. "Variability is natural, predictable, and quantifiable" had the greatest frequency of evidence in Cases A and B, the ones with mathematically able students, and almost non-existent evidence in Case C. The importance of random sampling and random assignment are other goals where the evidence in Cases A and B exceed that of C. Looking ahead to the second block of goals, Cases A and B also make a distinction between the mathematical and everyday meanings of "random" that is ignored by Case C.

Case C matches the well-aligned Case B regarding the goal "data beat anecdotes." Both cases begin the semester with discussions of the importance of data in understanding a topic of interest and make a connection between statistical inference and the scientific method (Case B Observation 1; Case C Observation 1). Case C also has equal evidence with Case B in the second block's goal of "how to determine when a cause-and-effect inference can be drawn from an association," though it is the weakest area of evidence in Case B.

Third and fourth blocks. These two blocks contain goals that are evident in all four cases. The goals listed here are where GAISE comes closest to suggesting a list of topics to be covered in an introductory course. The third block might be thought of as procedural while the fourth focuses on conceptual understanding of inference.

Every goal in the third block references "how to …" do something with data or statistical results. Several of these goals are noticeable at first glance—in the course syllabi. The goal "how to obtain or generate data" is evident in Case A's schedule where it mentions "Simulation," and Case B's topic list includes "Surveys and designed experiments." Case A devotes a class to "Summary Statistics and Graphical Summaries" that covers the goals "how to graph the data" and "how to interpret numerical summaries and graphical displays." Case C has a lab for "Frequency Analysis" as well as "Central Tendency and Variability" that address the same goals.

In all four cases, evidence that they share the goal that students should know "how to make appropriate use of statistical inference" is plentiful. Cases C and D spend entire class sessions on activities that give students practice in choosing an appropriate inferential procedure (Case C Observation 6; Case D Observation 8). Case A introduces two sample t-tests with emphasis on the different conditions that dictate different procedures (Case A Observation 6). Case B is explicit about the importance of random sampling as the basis for the procedures in the course (e.g., Case B Lecture Topic 6). "Communicating the results of a statistical analysis" is evident in all four cases as well. Careful statements of both a statistical conclusion and a contextualized one are explicitly demanded by the various instructors (e.g., Case A6_Hypothesis Testing, Case B Lecture Topic 6, Case C Lecture 15, Case D Lecture 7).

The concepts of inference in the fourth block of goals have universal alignment across the cases. Case C ran out of time to cover confidence intervals during the semester or it may have been unanimously well-aligned. Each instructor began the semester with expectations of stressing the concepts of inference (e.g., Case A Syllabus

and Case B Interview 1). They all discussed the application of sampling distributions and statistical significance with attention to the particular vocabulary mentioned by GAISE. With the already noted exception of Case C, confidence intervals and the identified vocabulary also received in-depth coverage by the professors.

Fifth block. The last block of goals is where these cases are least aligned with GAISE. All cases have multiple sources of evidence that students "should know how to interpret statistical results in context" but have little or no evidence regarding the "ability to critique news stories and journal articles" or "when to call for help from a statistician." All the cases mention critiquing stories and articles but do not offer opportunities for practice or assess the students' ability to do so. They all suffer from a lack of explicit discussion of times when more complicated statistical procedures necessitate reference to a statistician.

Variety in Enacting Recommendations for Teaching

All four cases demonstrate alignment with the GAISE recommendations to "emphasize statistical literacy and develop statistical thinking" and "stress conceptual understanding, not merely knowledge of procedures." None of the four cases aligns with the recommendation to "use real data." The other three recommendations have mixed alignment among the cases (see Table 27).

Table 27				
Matrix of Recommendations for Teaching				
	Case A	**Case B**	**Case C**	**Case D**
Emphasize statistical literacy and develop statistical thinking	well-aligned	aligned	well-aligned	aligned
Use real data	potential	potential	uneven	potential
Stress conceptual understanding, not merely knowledge of procedures	aligned	aligned	aligned	aligned
Foster active learning in the classroom	uneven	potential	well-aligned	potential
Use technology for developing concepts and analyzing data	potential	well-aligned	aligned	aligned
Use assessments to improve and evaluate student learning	uneven	potential	aligned	potential

The unanimous efforts of these four instructors to emphasize statistical literacy and stress conceptual understanding speaks to the success of Cobb's chapter on statistics education in the 1992 MAA Notes, *Heeding the Call for Change*. Using technology for developing concepts and analyzing data was evident in three of the four cases and not completely neglected in the fourth. The other facets of Cobb's recommendations (use real data and foster active learning) incorporated into GAISE are less evident in these four cases. The only completely new recommendation in GAISE, the use of "assessments to improve and evaluate student learning," is challenging for most of these instructors.

Professor A is the least dependent on technology in the administration of the course. The addition of clicker quizzes in the course improved active learning and the use of assessments for learning but, unfortunately, did not move the case toward using technology as a tool for developing concepts. The instructor's enthusiasm for the re-designed course in spite of hurdles faced during the semester leaves open the possibility

of further progress in Case A's two areas of uneven alignment to GAISE and may encourage later introduction of technology for developing concepts.

Students in all cases had the opportunity to use technology for analyzing data. Cases C and D compared graphs created by the instructors using SPSS to develop concepts such as skewness and correlation. Case B used dynamic Excel demonstrations (see Appendix D) to illustrate repeated sampling, testing, construction of confidence intervals and the connections between these concepts. Furthermore, the Case B demonstrations link data, graphs, and numerical analyses to help students solidify their understanding through multiple representations.

All of the instructors struggled with using real data. Professor C collected data from the students early in the semester and used it on occasion in class, the only instructor taking this approach suggested by GAISE. Like the other professors, however, other work in class and lab used data sets whose origins were unknown to students. Cases A, B, and D used textbooks that specify that they encourage the use of real data and offer data sets on the accompanying CD or companion website. The textbook, therefore, contains the potential for implementation of this long-standing recommendation. It may even be true that the professors are already using real data without acknowledging that to the students or providing opportunities for the students to work with it themselves.

Case A differs from Cases B and D on both active learning and use of assessment for learning by the use of clicker quizzes in most class sessions. It is also a benefit to Case A's students that complete solutions to homework, written quizzes, and exams are available through the course management system and frequent think-pair-share activities

in class took place during the second half of the semester. Case B offers students correct answers for graded assignments but they come with no explanations. Professor B encourages students to work together during recitations but there is no imperative to do so and students were observed waiting for the instructor's solution. The feedback that TAs give students in Case D was not observed nor were the activities in lab to provide evidence in favor of these two recommendations. Case C—the only one aligned with GAISE in either of these two areas—used in-class activities regularly, took student questions and input during lectures as well as lab, and provided precise, hand written feedback on exams.

Products of the Patten-Matching Analyses

The cross-case analyses brought to light four themes related to the ways that the diverse cases in this study do and do not implement the American Statistical Association's *Guidelines for Assessment and Instruction in Statistics Education* (2005):

Theme 1 - Know thy students

Theme 2 - Small changes, big differences

Theme 3 - Procedures *and* concepts

Theme 4 - Statistical literacy for critiquing claims

One additional theme emerged that did not directly relate to the research questions motivating this study but, nonetheless, colored the case descriptions and the subsequent analyses: awareness of GAISE is not required for implementation of its goals and recommendations. Further discussion of this theme is not necessary here but the unanimous instructor unfamiliarity with GAISE should be kept in mind when considering its implementation in the courses participating in this study.

153

Theme 1 - Know thy students. The professors participating in this study knew the predispositions of the student population that would fill their classes. In their initial interview, they each talked about the mathematical preparation—or lack thereof—and motivations of their students before they ever met them. The textbooks they selected match their students in aptitude and attitude as well as aiming at disciplinary relevance. Software selection is similarly appropriate to the careers available from the chosen major.

They are also forward-thinking, knowing that students will use the material in the future. Professor D's response in the final interview represents the other instructors' thoughts on their students' future with statistics: "They go get a job and discover that they have to organize some data or run a small analysis. That's when they discover that the topic they had no use for in college is useful in their career." All four professors express confidence that *all* of their students have gained useful skepticism as consumers of statistics regardless of their success as producers.

The GAISE College Report likens introductory statistics courses with a focus on statistical literacy and being consumers of data to an art appreciation course, while courses focused on producing statistical analyses more closely resemble a studio art course. "Most courses are a blend of consumer and producer components, but the balance of that mix will determine the importance of each recommendation we present" (ASA, 2005, p. 11). The varying degrees of evidence within the cases in this study illustrate the spectrum described. The awareness these professors have for the preparation and expectations that their students arrive with, as well as the career paths the students are on appropriately influence much of the content selected for these courses.

154

Theme 2 - Small changes, big differences. Three of the four professors introduced some kind of change into their courses during their participation, two of them specifying the intention of improving an area that GAISE recommended as important pedagogy. Case A had several small changes to lectures (less passive listening, more active doing) and assessments (formal and informal use of clicker responses) that represent a large paradigm shift (students held responsible for reading the text *before* the lecture) for the professor and for some students. Professor B attended and participated in recitation sessions, which doubled the weekly contact hours with students. Online homework provided students in Case C more immediate feedback on their understanding as well as additional opportunities to check their conceptual understanding. The casual implementation of clickers for a couple of activities early in the semester also occurred in Case C.

Evaluating the success of these changes is not the intention of this study but they contributed to the evidence of the cases' implementation of GAISE teaching recommendations. Without the use of clickers in Case A, evidence of active learning and assessment to improve learning would have been far weaker. In the final interview, Professor A expressed the intention to continue using the clickers and identified areas where their use could be increased in future semesters. Further experience incorporating this one change has potential for bringing Case A into alignment with GAISE without additional restructuring of the course. Professor C also reflected on the positive impact that the online homework system brought to the course through improved homework grades (multiple attempts to achieve correct answers) and the inclusion of conceptual questions that did not appear in the instructor-generated homework of previous semesters.

Both of these effects contributed evidence of GAISE alignment regarding assessment for learning. The addition of the clicker activities added to the already sufficient evidence of active learning in Case C. Professors B and D may wish to consider the incorporation of clickers or addition/enhancement of online homework into their own attempts at improving their pedagogy.

Theme 3 - Procedures *and* concepts. The universal evidence for alignment between the cases and the third and fourth blocks of GAISE goals speaks to the progress of reform in statistics education begun by Cobb's 1992 report. These professors are committed to ensuring that students understand the procedures they carry out, knowing the *why* and the *when* as well as the *what* and the *how*. Without further investigation, it is impossible to say whether the instructor's intention is the cause of the textbook selection or the effect of textbook authors/publishers following first Cobb and then GAISE recommendations. In either case, none of the professors in this study was content to simply present a menu of statistical analyses or dwell on theoretical statistics. The depth of explanation for the mathematical operations within each procedure varies across the cases (coincidentally, descending in alphabetical order) but the emphasis on conceptual understanding and when a particular procedure is appropriate remained uniform.

Theme 4 - Statistical literacy for critiquing claims. Cases A and D include critical thinking about statistical claims in the course objectives listed on the syllabus; while Professors B and C are less formal, they do mention it as a goal for the course during the initial interviews. In the final interviews, they all expressed confidence that students had learned to be critical of published statistics; however, none had assessed that

ability. GAISE begins its recommendation regarding assessment with the statement "students will value what you assess" (p. 13) that calls into question the professors' commitment to this objective for the course.

At some point in each case, students were encouraged to consider how statistics might be misleading either out of ignorance or by intention. These instances took place piecemeal, as a topic that could be misused was covered (e.g., sampling bias when discussing random samples or causal claims when discussing correlation). The professors did not model a general critique of either popular media reports or professional journal articles. Case C's textbook demonstrates the critical thinking that the professor wants students to adopt but is not discussed in any observed class.

Chapter 6: Discussion

The GAISE College Report offers "a list of goals for students, based on what it means to be statistically literate" and "recommendations regarding the need to focus instruction and assessment on the important concepts that underlie statistical reasoning" (ASA, 2005, p. 1). It is against these goals and recommendations that this study has compared the four cases—both individually and collectively. The detailed descriptions provide answers to the two research questions motivating this study:

- How do the introductory statistics courses offered by different academic departments define objectives and deliver instruction?

- Are there sufficient commonalities for students in *all* classes to achieve the level of statistical literacy and thinking recommended by the GAISE College Report?

The detailed descriptions of the four case studies in chapter four in conjunction with the comparisons of the structural compositions that begin the cross-case analysis in chapter five answer the question of how courses differ across disciplines. Although there is little discussion of the first research question here, reference to these similarities and differences are inevitable in discussing the sufficiency of the cases' alignment with GAISE. Reference to Table 25 (p. 141) may be useful for the reader.

Answering the second research question is a more complex endeavor. The cross-case analysis in chapter five focuses on the alignment of the cases to the goals and recommendations of GAISE. Further discussion of the themes from that analysis and their implications for statistics education research will comprise the bulk of this chapter.

Mention of the limitations inherent in this study and plans for future research will conclude the chapter and the report.

Statistically Educated Students

It bears repeating that the GAISE report is predicated on the idea that the "desired result of all introductory statistics courses is to produce statistically educated students, which means that students should develop statistical literacy and the ability to think statistically" (ASA, 2005, p. 11). The word "all" is what this study's research questions examine. Keeping in mind the descriptions of how these four courses are implemented, attention to the commonalities across the cases will answer the question of sufficient opportunities for students in different disciplines to gain statistical literacy and develop statistical thinking.

Themes. The cross-case analysis of chapter five results in four themes related to the ways that the GAISE goals and recommendations are evident among the cases. The first two reflect instructor interest in student success, while the final two reveal what the instructors envision as success for their students.

Interest in student success. The professors participating in this study have a deep understanding of both their students and their subject. Lee Shulman (1988) would describe this as pedagogical content knowledge: "The teacher not only understands the content to be learned and understands it deeply, but comprehends which aspects of the content are crucial for *future* understanding of the subject and which are more peripheral and are less likely to impede future learning if not fully grasped" (p. 2). Selection of the textbook, organization of lectures, inclusion of technological tools, presentation of tasks for students (formally assessed or not), and final assignment of course grades are all

affected by the instructor's pedagogical content knowledge. The cases in this study represent four different disciplines with diverse statistical praxis, which is evident in the breadth and depth of the content included in the courses. This variety within the content coverage did not affect the cases' alignment with the GAISE goals.

In a strictly pedagogical sense, these professors show concern for providing the best possible environment for student learning. Each instructor expressed interest in offering an active learning environment, appreciation of the usefulness of statistical software, and a desire for authentic assessment. These same three areas arose during the final interviews while reflecting on what went well (or did not) during the observed semester. Though the implementation of these ideas manifested in varying ways across the cases, it is evident that three of the six GAISE recommendations for teaching are already part of the instructors' pedagogy.

Three of the instructors mentioned class size as a hindrance to active learning. The GAISE report offers some suggestions—both general and specific—for implementing projects and activities in large classes that the instructors might consider now that they are aware of this resource. Similarly, class size influences the types of assessment used in these courses and GAISE suggestions may be useful in the three cases that lacked evidence of using assessment for student learning. The one case that did not align with the recommendation for using technology to develop concepts and analyze data had introduced some technology regarding assessments, which may indicate willingness to consider further inclusion of software or web applications in lectures.

There was minimal evidence of alignment to the GAISE recommendation to use real data in any of the cases. Case C was the only one to collect data directly from the

students but using that data was rarely observed during this study. All cases have textbooks that include real data that could be used. The explicit awareness of its importance for student learning and the already available data make this an area easily improved in these courses.

What student success looks like. Invariably the cases emphasized the importance of conceptual understanding of statistical procedures in the written and verbal evidence collected. It is disappointing to find that formal assessments are so often focused on procedural skill. Interpretations of inferential results, however, provide the balance between conceptual understanding and knowledge of procedures that the professors endorse in agreement with GAISE. The unanimous alignment with the third and fourth blocks of goals reflects the interest in students' ability to perform procedures (with computational support), draw appropriate conclusions from the results, and communicate those conclusions to answer questions.

The importance of statistical literacy and thinking are likewise emphasized by the professors on syllabi and in interviews. Lectures cover both the "language of statistics" and the "fundamental ideas of statistics" (ASA, 2005, p. 14) though assessment of student literacy is mainly implicit through tasks that require selection of a procedure or interpretation of a result. Statistical thinking, however, is discussed, modeled, and assessed piecemeal rather than "solving statistical problems from conception to conclusion" (ASA, 2005, p.15). The individual and cross-case analyses gather this piecemeal treatment as evidence of alignment with the teaching recommendation and some of the goals in multiple blocks (e.g., association is not causation from the first block and how to interpret statistical results in context from the fifth block). Every professor

161

agreed with the GAISE goal for students to learn "how to critique news stories and journal articles that include statistical information, including identifying what's missing in the presentation and the flaws in the studies or methods used to generate the information" (ASA, 2005, p.13) but none ever demonstrated a critique to students or provided an opportunity for students to do so themselves.

Sufficient? Statistics education researchers already address questions of student outcomes in courses with and without GAISE-inspired instruction (see Chapter 2 for a review of that literature). The purpose of this study is not to evaluate the effectiveness of instruction aligned with GAISE but to discover if the goals for students as well as recommendations for teaching apply to courses taught in various disciplines. The pattern-matching strategy of this study shows that each of these cases shared many of the goals for students listed by GAISE, though the strength and variety of evidence found in the individual cases is not distributed in the same way. Other than the use of real data, the instructors acknowledge the importance of the teaching recommendations from GAISE. This noteworthy agreement comes without the instructors' knowledge of GAISE before their participation in this study.

The cases demonstrate that statistical literacy is important in all four disciplines. Less certain is their interest in developing statistical thinking, particularly in the ability to critique published statistics. Instructors expect their students to translate their skill as producers of statistics into being critical consumers of statistics with no assessment of their success in doing so. This gap in alignment is crucial to the overall goal of producing statistically educated citizens and needs further investigation of student ability to meet the instructors' expectation.

Are there sufficient commonalities for students in all classes to achieve the level of statistical literacy and thinking recommended by the GAISE College Report? These cases show that the disciplinary situation does not impact the ability of courses to meet the guidelines endorsed by the ASA. The variability in content among the courses still covers the "fundamental ideas" (ASA, 2005, p. 14) that should lead to the "desired result of all introductory statistics courses" (p. 11) for statistically educated students. The non-perfect alignment to the goals and recommendations of GAISE are not widespread enough in any one case to suspect that students leave the course without having gained *some* statistical literacy as the instructors aver. If there is a cause for concern, it is in the area of being critical consumers of statistics since it is never assessed. This concern applies across these disciplines.

Limitations of the Study

The conclusion just drawn, of course, comes with some cautions. The usual concerns about researcher bias, missed data, and misinterpretation of implicit intentions are reasonable points of discussion. Peer review and member checking in addition to the researcher's awareness of these concerns are attempts to minimize these issues. The question of missed data applies specifically to three areas: the courses not included in the study, the interactions that teaching assistants had with students, and the student perspectives on course implementation.

There is, perhaps, some unclaimed value in observing courses where the instructor is teaching the material for the first time or otherwise reluctant to be observed. The struggle to find a balance between content coverage and student understanding that faces a novice instructor could provide some interesting perspective on how the

163

experienced instructors came to include so many of the GAISE goals and recommendations in their courses without awareness of the guidelines.

It would be naïve to expect that no teaching and learning occurs when teaching assistants connect with students. In this study, the instructors without awareness of GAISE may still have passed on ideas of good teaching to their assistants or, perhaps, the TAs are aware of the guidelines from their own interest in educational research. These thoughts call to mind the report from Green's (2010) work with TAs at the University of Nebraska-Lincoln (see Chapter 2).

All of the evidence considered in this study comes from the instructor's perspective. Even the most explicit intention may be misinterpreted by students. Consideration of the student perspective would strengthen evidence where instructor intentions and actions align with GAISE or provide points of reflection where alignment is missing or illusory.

Suggestions for Future Research

The three specific limitations discussed above should be addressed in any follow-up studies that might arise. Reflections from these four professors when they next teach these courses could prove interesting now that they have gained awareness of GAISE. Observation of a course designed for health science students or graduate students in the professional schools would provide a more complete understanding of how diverse the "family of courses" is.

Existing statistics education research that evaluates student outcomes has focused on GAISE's teaching recommendations with little or no reference to the goals. Mapping the available tools for assessing statistical literacy and thinking to the GAISE goals may

be useful in bringing large data sets to the aid of curriculum designers and individual instructors. Such research may also be useful to the on-going dialogue regarding second courses.

Some of the emergent coding from this study suggests research topics that are not directly related to GAISE. Research/data ethics is not related to any of the goals for students but is included in two of the cases. There may be interesting connections between ethics instruction and student ability to critique statistical claims. All of the textbooks mentioned some important contributors to the discipline but the instructors did not. There is need for research on the usefulness of historical connections on student learning and attitudes toward statistics. The use of technology as administrative support and the role of teaching assistants in an introductory course are research topics that extend beyond statistics.

References

American Statistical Association (2005). *Guidelines for assessment and instruction in statistics education: College report*. Retrieved from http://www.amstat.org/education/gaise/

Ben-Zvi, D., & Garfield, J. (2008). Introducing the emerging discipline of statistics education. *School Science & Mathematics, 108*(8), 355-361. doi:10.1111/j.1949-8594.2008.tb17850.x

Bookman, Ganter, & Morgan (2008). Developing assessment methodologies for quantitative literacy: A formative study. *The American Mathematician, 115*(10), 911-929.

Berenson, M. L., Utts, J., Kinard, K. A., Rumsey, D. J., Jones, A., & Gaines, L. M. (2008). Assessing student retention of essential statistical ideas: Perspectives, priorities, and possibilities. *American Statistician, 62*(1), 54-61. doi:10.1198/000313008X272761

Burke, B., C., M., & Carnegie Foundation for the Advancement of Teaching, M. (2007, October 3). *A Mathematician's Proposal. Carnegie Perspectives*. Carnegie Foundation for the Advancement of Teaching. (ERIC Document Reproduction Service No. ED498954)

Campbell, A., Kirsch, I.S., and Kolstad, A. (1992). *Assessing Literacy: The Framework for the National Adult Literacy Survey*. Washington, D.C.: National Center for Education Statistics.

Chance, B., & Rossman, A. (2001). Sequencing topics in introducing statistics: A debate on what to teach when. *The American Statistician. 55*, 140-144.

Chiesi, F., & Primi, C. (2010). Cognitive and non-cognitive factors related to students' statistics achievement. *Statistics Education Research Journal, 9*(1), 6-26.

Chiou, C. (2009). Effects of concept mapping strategy on learning performance in business and economics statistics. *Teaching in Higher Education, 14*(1), 55-69.

Cobb, G. (1992). Teaching statistics. In Lynn A. Steen (Ed.), *Heeding the call for change: Suggestions for curricular action* (MAA Notes No. 22), 3-43.

David, F. N. (1970). Dicing and Gaming (A Note on the History of Probability). In E.S. Pearson and M.G. Kendall (Eds.), *Studies in the History of Statistics and Probability* (pp. 1-14). London: Charles Griffin & Company Limited.

delMas, R.C., Garfield, J. & Chance, B.L. (1999). A model of classroom research in action: developing simulation activities to improve students' statistical reasoning. *Journal of Statistics Education*, *7*(3).

Delcham, H., & Sezer, R. (2010). Write-skewed: Writing in an introductory statistics course. *Education, 130*(4), 603-615.

Department of Statistical Science (2008). Department History. Retrieved from http://www.ucl.ac.uk/statistics/department/history

Department of Statistics (n.d.). About us. Retrieved from http://www.stat.ncsu.edu/about.php

DeVeaux, R.D., Velleman, P.F., & Bock, D.E. (2006). *Intro Stats* (2nd Ed.). Boston,

MA: Pearson Education, Inc.

Dossey, J. A. (1997). National indicators of quantitative literacy. In L. A. Steen (Ed.),

Why numbers count: Quantitative literacy for tomorrow's America (pp. 48-54). New

York, NY: The College Board.

Ellis, W. (2001). Numerical common sense for all. In L. A. Steen (Ed.), *Mathematics and*

democracy: The case for quantitative literacy (pp. 61-65). National Council on

Education and the Disciplines.

Everson, M., Zieffler, A., & Garfield, J. (2008). Implementing new reform guidelines in

teaching introductory college statistics courses. *Teaching Statistics, 30*(3), 66-70.

doi:10.1111/j.1467-9639.2008.00331.x

Fink, A., & Lunsford, M. (2009). Bridging the divides: Using a collaborative honors

research experience to link academic learning to civic issues. *Honors in Practice*, *5*,

97-113.

Folks, J.L. (2002). The fragile beginnings at Oklahoma State University. Retrieved from

http://statistics.okstate.edu/information/history/fragile_beginnings.htm

Gabrosek, J., Stephenson, P., & Richardson, M. (2008). How LO can you GO?.

Mathematics Teacher, 101(7), 545-549.

Gal, I. (2002). Adults' Statistical Literacy: Meanings, Components, Responsibilities.

International Statistical Review, *70*, 1-25.

Garfield, J. (1995). How students learn statistics. *International Statistical Review, 63*, 25–

34.

Garfield, J. (2000). *An evaluation of the impact of statistics reform: Final Report.* National Science Foundation (REC-9732404).

Garfield, J., & del Mas, R. (2010). A web site that provides resources for assessing students' statistical literacy, reasoning and thinking. *Teaching Statistics, 32*(1), 2-7. doi:10.1111/j.1467-9639.2009.00373.x

Garfield, J., & Everson, M. (2009). Preparing teachers of statistics: A graduate course for future teachers. *Journal of Statistics Education, 17*(2).

Green, J. L. (2010). Teaching highs and lows: exploring university teaching assistants' experiences. *Statistics Education Research Journal, 9*(2), 108-122.

Gillman, R. (Ed.). (2006). *Current Practices in Quantitative Literacy*. Washington, D.C.: Mathematical Association of America.

Gnaldi, M. (2006).The relationship between poor numerical abilities and subsequent difficulty in accumulating statistical knowledge. *Teaching Statistics, 28*, 49-53.

Hall, M. R., & Rowell, G. (2008). Introductory statistics education and the national science foundation. *Journal of Statistics Education, 16*(2).

Hassad, R. A. (2008). Reform-oriented teaching of introductory statistics in the health, social and behavioral sciences - historical context and rationale. Proceedings of World Academy of Science: Engineering & Technology, 42398-403.

Heyde, C.C. & Seneta, E. (Eds.). (2001). *Statisticians of the Centuries*. New York: Springer-Verlag.

Hobbs, J. (2008). Leaders and landmarks: 75 years of statistics at Iowa State University (1933-2008). Retrieved from http://www.public.iastate.edu/~jonhobbs/Posters/StatHistory.pdf

Holt, M., & Scariano, S. M. (2009). Mean, median and mode from a decision perspective. *Journal of Statistics Education, 17*(3).

Kaplan, J.J., Fisher, D. & Rogness, N. (2009). Lexical Ambiguity in Statistics: What do students know about the words association, average, confidence, random and spread? *Journal of Statistics Education*, *17*(3), http://www.amstat.org/publications/jse/v17n3/kaplan.pdf

Kirsch, I.S. & Jungblut, A. (1986). *Literacy: Profiles of America's young adults.* Princeton, NJ: Educational Testing Service.

Kirsch, I.S., Jungblut, A., Jenkins, L., & Kolstad, A. (1993). *Adult literacy in America: A first look at the findings of the national adult literacy survey.* Princeton, NJ: Educational Testing Service.

Krefting, L. (1999). Rigor in qualitative research: The assessment of trustworthiness. In in A. Milinki (Ed.). *Cases in Qualitative Research*. Los Angeles, CA: Pyrczak Publishing, 173-181.

Lesser, L. M., & Winsor, M. S. (2009). English language learners in introductory statistics: lessons learned from an exploratory case study of two pre-service teachers. *Statistics Education Research Journal, 8*(2), 5-32.

Lutzer, D.J., Rodi, S.B., Kirkman, E.E., & Maxwell, J.W. (2007). Statistical abstract of undergraduate programs in the mathematical sciences in the United States: Fall 2005 CBMS survey. American Mathematical Association. Retrieved from http://www.ams.org/profession/data/cbms-survey/cbms2005

Madison, B. (2004). Two mathematics. *Peer Review*, *6*(4), 9-12.

Marshall, C. & Rossman, G.B. (1999). *Designing qualitative research* (3rd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Mason, R.L. (1999). ASA: The first 160 years. Retrieved from http://www.amstat.org/about/first160years.cfm

McClure, R., & Sircar, S. (2008). Quantitative literacy for undergraduate business students in the 21st century. *Journal of Education for Business*, *July/August*, 369-374.

Merritt, E.G., Rimm-Kaufman, S.E., Berry, R.Q., Walkowiak, T.A., & McCracken, E.R. (2010). A reflection framework for teaching math. *Teaching Children Mathematics, 17*(4), 238-248.

Meyer, J. & Dwyer, C. (2005). Improving quantitative reasoning through analysis of news stories. *International Journal of Learning*, *12*(6), 165-173.

Miles, M.B. & Huberman (1994). *Qualitative data analysis* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Moore, D.S., McCabe, G.P., Duckworth, W.M., & Alwan, L.C. (2009). *The practice of business statistics: Using data for decisions* (2nd ed.). New York, NY: W.H. Freeman and Company.

National Center for Education Statistics (2009). *Digest of education statistics, 2008*. Retrieved from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009020

National Council of Teachers of Mathematics (2000). *Principles and Standards for School Mathematics*. Reston, VA: NCTM.

National Council of Teachers of Mathematics (2007). *Mathematics Teaching Today: Improving Practice, Improving Student Learning*. Reston, VA: NCTM.

Navidi, W.C. (2011). *Statistics for Engineers and Scientists* (3rd ed.). Boston, MA: McGraw-Hill.

Nelson, D. (2009). Using simple linear regression to assess the success of the Montreal Protocol in reducing atmospheric chlorofluorocarbons. *Journal of Statistics Education, 17*(2).

Neumann, D. L., & Hood, M. (2009). The effects of using a wiki on student engagement and learning of report writing skills in a university statistics course. *Australasian Journal of Educational Technology, 25*(3), 382-398.

O'Connor, J.J. & Robertson, E.F. (Eds.) (n.d.). The MacTutor history of mathematics archive. Retrieved from http://www-history.mcs.st-andrews.ac.uk/history/index.html

Pagano, R.R. (2010). *Understanding Statistics in the Behavioral Sciences* (9th ed.). Belmont, CA: Wadsworth, Cengage Learning.

Pearson, E.S. (1936). Karl Pearson: An appreciation of some aspects of his life and work. *Biometrika, 28*(3/4), 193-257.

Petocz, P. & Reid, A. (2003). Relationships between students' experience of learning statistics and teaching statistics. *Statistics Education Research Journal, 2*(1), 39–53.

Petocz, P., & Reid, A. (2005). Rethinking the tertiary mathematics curriculum. *Cambridge Journal of Education, 35*(1), 89-106. doi:10.1080/03057640420003332515

Pfannkuch, M., Regan, M., Wild, C., & Horton, N. J. (2010). Telling data stories: Essential dialogues for comparative reasoning. *Journal of Statistics Education, 18*(1).

Phelps, A. L., & Dostilio, L. (2008). Studying student benefits of assigning a service-learning project compared to a traditional final project in a business statistics class. *Journal of Statistics Education, 16*(3),

Richardson, M., Stephenson, P., & Gabrosek, J. (2010). How LO can you GO?. *Teaching Statistics, 32*(1), 8-12. doi:10.1111/j.1467-9639.2009.00389.x

Root, R. (2009). Social justice through quantitative literacy: A course connecting numeracy, engaged citizenship, and a just society. *Democracy & Education*, *18*(3), 37-43.

Rossman, A. J. (2008). Reasoning about informal statistical inference: One statistician's view. *Statistics Education Research Journal, 7*(2), 5-19.

Royal Statistical Association (n.d.). History. Retrieved from http://www.rss.org.uk/site/cms/contentCategoryView.asp?category=42

Rumsey, D.J. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education, 10*(3).

Sanders, L. (2011). Antidepressants show signs of countering Alzheimer's. *Science News, 180*(6), 5-6.

Scheaffer, R.L. (2003). Statistics and quantitative literacy. In B.L. Madison & L.A. Steen (Eds.), *Quantitative literacy: Why numeracy matters for schools and colleges* (pp. 145-152). Princeton, NJ: National Council on Education and the Disciplines.

Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (pp. 957-1009). Charlotte, NC: Information Age Publishing.

Shulman, L.S. (1988). A union of insufficiencies: Strategies for teacher assessment in a period of educational reform. *Educational Leadership,* Nov, 36-41.

Sisto, M. (2009). Can you explain that in plain English? Making statistics group projects
    work in a multicultural setting. Journal of Statistics Education, 17(2).

Spence, I. & Wainer, H. (1997). Who was Playfair? *Chance*, *10*(1), 35-37.

Stake, R. (2000). Case Studies. In N. Denizen & Y. Lincoln (Eds.)*, Handbook of
    qualitative research* (2nd ed., pp. 435-454). Thousand Oaks, CA: Sage Publications,
    Inc.

Steen, L.A. (Ed.). (1997). *Why numbers count: Quantitative literacy for tomorrow's
    America*. New York: College Entrance Examination Board.

Steen, L.A. (Ed.). (2001). *Mathematics and democracy: The case for quantitative
    literacy*. The National Council on Education and the Disciplines.

Steen, L.A. (Ed.). (2004). *Achieving quantitative literacy: An urgent challenge for higher
    education*. The Mathematical Association of America.

Stigler, S. M. (1986). *The history of statistics: The measurement of uncertainty before
    1900*. Cambridge, Massachusetts: The Belknap Press of Harvard University Press.

Theoret, J. M. & Luna, A. (2009). Thinking statistically in writing: Journals and
    discussion boards in an introductory statistics course. *International Journal of
    Teaching & Learning in Higher Education, 21*(1), 57-65.

Tolbert-Bynum. (2008). *Research Starters: Adult Literacy Programs.* EBSCO
    Publishing Inc.

Trochim, W.M.K. (1989). Outcome pattern matching and program theory. *Evaluation
    and Program Planning*, 12, 355-366.

Utts, J. (2003). What educated citizens should know about statistics and probability.
    American Statistician, *57*(2), 74–79.

Wild, C.J. & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry.

   International Statistical Review. 67(3), 223–265.

Wilks, S.S. (1950). Undergraduate statistical education. *Journal of the American*

   Statistical Association, 46(253), 1-18.

Wade, B. & Goodfellow, M. (2009). Confronting statistical literacy in the undergraduate

   social science curriculum. *Sociological Viewpoints*, Fall2009, 75-90.

Wiest, L.R., Higgins, H.J., & Frost, J.H. (2007). Quantitative literacy for social justice.

   Equity & Excellence in Education, 40: 47-55.

Yin, R.K. (1994). *Case study research: Design and methods* (2nd ed.). Thousand Oaks,

    CA: Sage Publications, Inc.

Zieffler, A. S., & Garfield, J. B. (2009). Modeling the growth of students' covariational

   reasoning during an introductory statistics course. *Statistics Education Research*

   Journal, 8(1), 7-31.

Appendix A: Observation Protocol

Observation Protocol

Case _____ Date _____ Recording made: Y   N

Students: _____Number of Males _____Number of Females_____Documents collected: Y   N

Observer's location within classroom: _____

1. According to the instructor/syllabus, the purpose of this lesson is:

2. The focus of this lesson is best described as: (Check one.)
   - ☐ Almost entirely working on the development of procedures/vocabulary
   - ☐ Mostly working on the development of procedures/vocabulary, but working on some statistical concepts
   - ☐ About equally working on procedures/vocabulary and working on statistical concepts
   - ☐ Mostly working on statistical concepts, but working on some procedures/vocabulary
   - ☐ Almost entirely working on statistical concepts
   - ☐ Administrative topics

3. *Instructional design* of the lesson as evident by instructor's verbal or written statement(s)

GAISE Recommendations	**Major Part**	**Part**	**Minor Part**	**Not Present**
Emphasize statistical literacy and develop statistical thinking				
Use real data				
Stress conceptual understanding, *not* merely knowledge of procedures				
Use technology for developing concepts and analyzing data				
Use assessments to improve and evaluate student learning				

4. *Implementation* of the lesson as evident by actual lecture/activity

GAISE Recommendations	Major Part	Part	Minor Part	Not Present
Emphasize statistical literacy and develop statistical thinking				
Use real data				
Stress conceptual understanding, *not* merely knowledge of procedures				
Use technology for developing concepts and analyzing data				
Use assessments to improve and evaluate student learning				

5. Classroom culture - does it "foster active learning"?

GAISE Recommendations	Major Part	Part	Minor Part	Not Present
Active participation of all was encouraged and valued.				
There was a climate of respect for student ideas, questions, and contributions.				
Interactions reflected collegial working relationships among students				
Interactions reflected collaborative working relationships between teacher and students.				
The climate of the lesson encouraged students to generate ideas, questions, conjectures, and/or propositions.				
Intellectual rigor, constructive criticism, and the challenging of ideas were evident.				

6. Content – topics (conceptual and procedural) work toward goals for students?

GAISE Goals for Students	Major Part	Part	Minor Part	Not Present
Students should believe and understand why…[1]				
Students should recognize…[2]				
Students should understand the parts of the process through which statistics works to answer questions.[3]				
Students should understand the basic ideas of statistical inference.[4]				
Finally, students should know…[5]				

[1] …why:
- Data beat anecdotes
- Variability is natural, predictable, and quantifiable
- Random sampling allows results of surveys and experiments to be extended to the population from which the sample was taken
- Random assignment in comparative experiments allows cause-and-effect conclusions to be drawn
- Association is not causation
- Statistical significance does not necessarily imply practical importance, especially for studies with large sample sizes
- Finding no statistically significant difference or relationship does not necessarily mean there is no difference or no relationship in the population, especially for studies with small sample sizes

[2] …recognize:
- Common sources of bias in surveys and experiments
- How to determine the population to which the results of statistical inference can be extended, if any, based on how the data were collected
- How to determine when a cause-and-effect inference can be drawn from an association based on how the data were collected (e.g., the design of the study)
- That words such as "normal," "random," and "correlation" have specific meanings in statistics that may differ from common usage

[3] namely:
- How to obtain or generate data
- How to graph the data as a first step in analyzing data, and how to know when that's enough to answer the question of interest
- How to interpret numerical summaries and graphical displays of data—both to answer questions and to check conditions (to use statistical procedures correctly)
- How to make appropriate use of statistical inference
- How to communicate the results of a statistical analysis

[4] including:
- The concept of a sampling distribution and how it applies to making statistical inferences based on samples of data (including the idea of standard error)
- The concept of statistical significance, including significance levels and p-values
- The concept of confidence interval, including the interpretation of confidence level and margin of error goals for students in an introductory course: what it means to be statistically educated

[5] …know:
- How to interpret statistical results in context
- How to critique news stories and journal articles that include statistical information, including identifying what's missing in the presentation and the flaws in the studies or methods used to generate the information
- When to call for help from a statistician

7.   This lesson was impacted by factors imposed on the instructor.

Influences	Positive	No Impact	Negative
Policy (university, department, academic calendar, etc.)			
Physical environment (presence & useability of technology, temperature, seating arrangement, etc.)			
Instructional materials (textbook, handouts, tools, etc.)			
Students (absenteeism, tardiness, disruptive behavior, etc.)			
Teacher (unwell, distracted, enthusiasm, current event, etc.)			

8. Overall "flavor" of the lesson with respect to GAISE recommendations and goals:
   - ☐ Well-aligned
   - ☐ Somewhat aligned
   - ☐ Some parts are aligned, others are not
   - ☐ Somewhat mis-aligned
   - ☐ Entirely mis-aligned

Narrative:

Appendix B: Interview Protocols

Semi-structured (pre-semester) Interview Protocol

<u>The course and the instructor</u>

How often is this course taught?

Has the content and/or teaching of this course changed over time?

How often do you teach this course?

How would you characterize your teaching? (SLrT, data, conceptual, active, technology, assess)

How has your teaching of this course evolved over time?

How is your teaching of this course similar or different from those who have previously taught the course?

How did you get assigned to teaching this course?

Do you look forward to teaching this course?

<u>Expectations</u>

What are some things students should know and be able to do prior to enrolling into this course?

Are most students able to do the things you described in the previous question?

Describe students who are successful in mastering the content of this course.

Are the students taking this course required to do so? If so, do you think they appreciate why?

After taking this course what should students know and be able to do?

Do you believe most students leave the course able to do those things?

Do you think students grow to appreciate the need for statistical education?

<u>Teaching</u>

Is there a topic/lesson that you find especially enjoyable to teach? Least enjoyable?

Is there a topic/content that is challenging for many students? Why? In way ways do you help students with this challenging topic?

Is there a topic/content that sparks student interest in statistical thinking?

Appendix C: Examples of Lecture Presentations

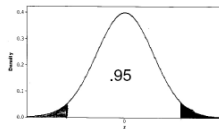Examples of lecture notes from Case A:

Section 1.2

Descriptive Statistics

How to describe/summarize a sample

6

Consider a sampling distribution of the mean from a normal population with $\sigma_{\overline{X}} = \sigma / \sqrt{n}$

$\Rightarrow$ $P(-1.96 \leq Z \leq 1.96) = .95$
 or
$P(\mu - 1.96\,\sigma_{\overline{X}} \leq \overline{X} \leq \mu + 1.96\sigma_{\overline{X}}) = .95$

.

$\therefore$ We expect $\overline{x}$ to be within $1.96\sigma_{\overline{X}}$ of μ 95% of time
Equivalently, μ is within $1.96\sigma_{\overline{X}}$ of $\overline{x}$ 95% of time

5.1   7

Linear Relationships

A linear relationship between the x and y values
is envisioned as a straight line through a scatterplot
(Reflects "direct" and "inverse" relationships)
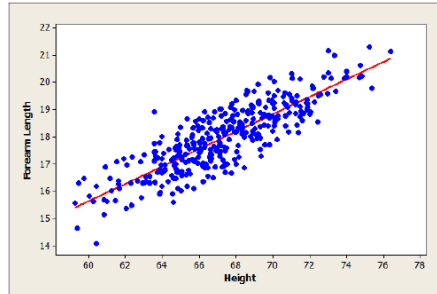
Examples of lecture notes from Case B:

Introduction to Inference
Estimating with Confidence

Section 6.1

What does is mean to accept or reject H_0?

- Accepting H_0 is a failure to observe sufficiently strong evidence against it.
  - "Accept H_0" ≠ "H_0 is true"

- Rejecting H_0 does not necessarily imply a "practically" significant departure from H_0.
  - Explore with graphical and descriptive statistics

- A conclusion about H_0 may be invalid if data production was poorly designed

Example: Wages and experience

Do wages rise with experience? In a study of employment trends, wage (y, in \$/week) and length of service (LOS = x, in months) measurements were obtained from $n = 59$ workers in similar customer-service positions.

Wages	LOS	Wages	LOS	Wages	LOS	Wages	LOS	Wages	LOS	Wages	LOS
389	94	403	76	443	222	486	60	547	228	443	104
395	48	378	48	353	58	393	7	347	27	566	34
329	102	348	61	349	41	311	22	328	48	461	184
295	20	488	30	499	153	316	57	327	7	436	156
377	60	391	108	322	16	384	78	320	74	321	25
479	78	541	61	408	43	360	36	404	204	221	43
315	45	312	10	393	96	369	83	443	24	547	36
316	39	418	68	277	98	529	66	261	13	362	60
324	20	417	54	649	150	270	47	417	30	415	102
307	65	516	24	272	124	332	97	450	95		

Examples of lecture notes from Case C:

Types of Graphs

X	f
12-13	4
10-11	5
8-9	3
6-7	3
4-5	2

184
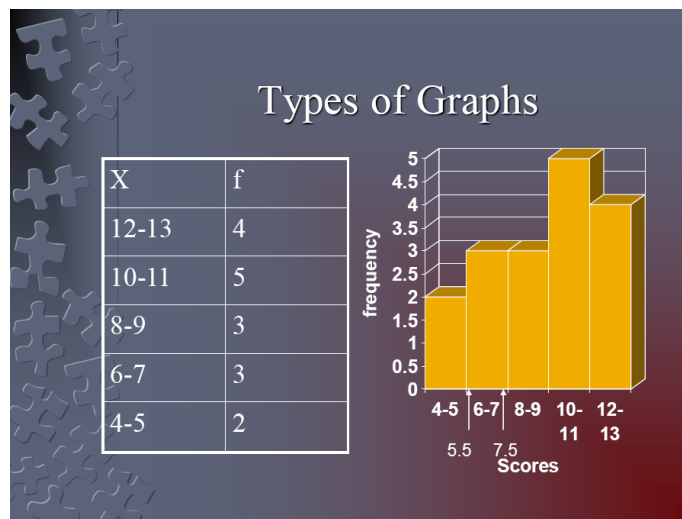
Additional Characteristics

- Positive vs. Negative relationships
  - <u>Positive relationships</u> – <u>direct</u> relationship between the variables
    - As X increases, Y increases
      - High scores on quiz 1 correspond to high scores on quiz 2
  - <u>Negative relationships</u> – <u>inverse</u> relationship between X and Y
    - As X increases, Y decreases
      - Number of clouds increase, brightness decreases

Teams!

- Create a scatterplot for the dataset. What kind of relationship is represented? How strong do you think this relationship is?
- What is the correlation for number of extra curricular activities and H.S. GPA?
  - How would you interpret this correlation?
- Determine the Least Squares Regression line for predicting H.S. GPA from the number of extra curricular activities.
  - How good is your model? In other words what proportion of variability for Y is predicted by X?
- Jane is involved in 3 clubs. What is your best prediction for her H.S. GPA?

Examples of lecture notes from Case D:

The Huffington Post

College Affordability: Damned If You Go, Damned If You Don't?

… Today, low-income students <u>must finance</u> an amount equivalent to 72 percent of their family's annual income to attend a public university for one year, even after accounting for grant aid. So is it any surprise that by age 24, you're <u>10 times more likely</u> to have a bachelor's degree if your parents are wealthy than if they're poor? Or that <u>for the first time</u> in our nation's history, student loan debt tops credit card debt?

Sampling Distribution of Means
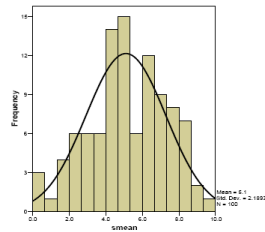
Correlation Coefficient

Correlations

		Hours studied	GPA
Hours studied	Pearson Correlation	1	.884**
	Sig. (2-tailed)	.	.000
	N	16	16
GPA	Pearson Correlation	.884**	1
	Sig. (2-tailed)	.000	.
	N	16	16

**. Correlation is significant at the 0.01 level (2-tailed).

10

Appendix D: Excel Demonstration

Population

Individual	Value
1	103
2	92
3	91
4	108
5	87
6	111
7	87
8	92
9	108
10	95
11	88
12	89
13	117
14	74
15	102
16	78
17	107
18	96
19	97
20	90
21	113
22	104
23	110
24	129
25	98
26	92
27	89
28	108
29	107
30	91
31	108
32	91
33	109
35	85
36	104
37	103
38	106
39	108
40	106
41	112
42	111
43	104
44	106
45	101
46	99
47	103
48	90
49	96
50	81
51	104
52	104

Repeated samples

Sampled — Sample

Individual	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	103	98	83	101	106	87	91	117	97	102	104	101	93	103	106	108	104	118	108
2	97	106	101	107	104	87	107	103	107	101	103	104	117	74	91	121	111	109	91
3	109	89	105	110	121	103	100	103	107	127	106	93	111	121	118	109	109	100	108
4	101	91	98	104	117	95	92	108	106	113	98	95	118	109	85	99	108	105	85
5	92	100	109	95	92	110	85	112	111	100	99	108	118	103	107	92	117	101	102
6	106	103	91	105	92	91	85	117	90	85	118	101	101	108	117	103	103	103	97
7	104	95	95	121	91	104	96	101	107	99	93	90	118	114	89	118	105	106	106
8	99	106	103	108	106	89	101	96	108	98	99	87	121	85	108	106	91	89	129
9	117	103	104	108	99	96	107	106	96	90	108	92	111	99	98	103	113	111	109
10	113	101	85	95	95	98	100	96	91	106	87	99	108	78	109	102	117	118	97
11	101	91	95	85	101	101	107	92	91	108	78	111	106	74	105	106	109	118	102
12	101	107	118	101	96	91	98	111	90	102	90	106	95	89	106	102	108	113	108
13	96	91	121	109	102	91	83	121	92	85	74	103	108	108	107	106	108	99	91
14	87	118	107	92	95	110	106	87	83	87	87	106	106	91	101	102	98	92	99
15	106	108	101	108	85	99	106	106	127	98	100	109	106	107	106	106	108	99	108
16	111	93	106	129	87	105	97	113	111	110	127	113	107	94	104	90	92	89	87
17	90	111	117	104	107	106	87	96	104	108	103	106	100	99	99	91	91	95	104
18	108	97	108	104	92	92	96	107	103	85	96	95	95	104	117	81	87	83	106
19	99	107	87	107	83	118	89	107	99	99	113	87	87	111	97	103	103	106	91
20	110	103	92	93	114	113	92	118	104	99	96	109	111	87	129	104	89	85	111
mean (of all 20):	102.5	100.8	101.2	104.2	99.3	99.3	96.3	105.8	101.3	99.6	99.7	100.7	106.9	97.8	105.0	102.7	103.6	101.8	102.0
two-sided P-val:	0.36	0.83	0.71	0.10	0.69	0.66	0.09	0.02	0.67	0.78	0.80	0.85	0.01	0.30	0.05	0.31	0.17	0.52	0.49
Reject H_a?	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
Lower CI bound:	97.8	96.1	96.5	99.5	94.7	94.6	91.6	101.1	96.6	95.0	95.0	96.1	102.2	93.2	100.3	98.1	98.9	97.1	97.3
Upper CI bound:	107.2	105.5	105.9	108.9	104.0	103.9	101.0	110.5	106.0	104.3	104.4	105.4	111.6	102.5	109.6	107.4	108.3	106.5	106.7
Cover 100.3?	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1

Sample size: n = 20
Population std. dev.: σ = 10.7
Std. dev. of x-bar: $\sigma_{x\text{-bar}} = \sigma/\sqrt{n}$ = 2.39

Significance level: α = 0.05

Claimed value of μ: μ_0 = 100.3
Interesting alternative value of μ: μ_a = 101.3

Margin of error $z^* \cdot \sigma/\sqrt{n}$ = 4.69   CI formula: x-bar ± 4.69

Percent of tests that "Reject H_a:" 5%
Percent of tests that "Accept H_a:" 95%
Significance level: 0.05

Reject H_0

Accept H_0

Percent CIs that do not cover μ: 5%
Percent CIs that cover μ: 95%
Confidence level: 0.95