THE GOAL OF INTRODUCTORY STATISTICS: REASONING ABOUT DATA

Milo Schield, Augsburg College
Dept. of Business & MIS. 2211 Riverside Drive. Minneapolis., MN 55454

KEY WORDS: Critical Thinking, Teaching, Induction, Quantitative Literacy, Statistical reform

ABSTRACT:
Given the continuing negative feedback from our students, it is appropriate to review the proper goals for the introductory course in applied statistics.

Section I summarizes some of the statements about the nature and goals of statistics. These statements appear to be relevant, factual and achievable. Yet if we were achieving these goals, students would find statistics valuable and useful while other faculty would see evidence of statistical reasoning by our students. All too often, these outcomes are not realized.

Section II examines some of the key terms in these statements about the nature and goals of statistics and finds them to be highly ambiguous. *Some teachers of statistics may not discern that these abstract goals may be interpreted narrowly or broadly, that these goals should be interpreted broadly and that teaching which achieves these narrow goals may not achieve these broader goals.*

Section III classifies various elements of statistics by method and subject matter. If some teachers were biased against topics involving induction, this might explain the failure of our students to internalize statistical thinking.

Section IV compares some of the important topics in statistics when goals are taken narrowly as compared to when goals are taken broadly. It also compares 'mathematical induction' with induction. Again, neither deduction nor 'mathematical induction' are wrong, but induction is also required.

Section V identifies some topics that are missing or under-emphasized.

In summary, teachers need to help students "read and interpret data".

I. GOALS IN TEACHING APPLIED STATISTICS

There are many excellent statements of the nature and goals of introductory applied statistics. Given the stature of the authors and the quality of the associated publication, the following statements have excellent credentials.

- "Our aim in the first course is to develop the critical reasoning skills necessary to understand our quantitative world. The focus of the course is the process of learning how to ask appropriate questions, how to collect data effectively, how to summarize and interpret that information, and how to understand the limits of statistical inference. Statistical thinking is central to education." [Robert Hogg (1990) in *Towards Lean and Lively Courses in Statistics* published by Gordon, Florence and Sheldon in *Statistics for the Twenty-First Century*]
- "Statistics is the science of data. More precisely, the subject matter of statistics is reasoning from uncertain empirical data." [David Moore (1992) in *Teaching Statistics as a Respectable Subject*]

Let us assume that these statements about the goals of statistics are fundamental and appropriate.

CONSEQUENCES
If we were achieving these goals, students
- would find the course useful and valuable
- might recommend statistics to others
- might keep their statistics book for reference
- might choose to take a second course

If we were achieving these goals, other teachers should see improvements among our students in their ability to
- summarize and display the data
- assess the quality of a study
- evaluate the applicability of a model
- evaluate the inferential conclusions
- communicate the nature and meaning of data
- assess the quality of a sample

If we were achieving these goals, our students would have internalized statistical thinking in ways that others could and would recognize.

PROBLEM
According to several studies, this is not happening. The failure to bring about these consequence indicates that statistics is not really achieving its stated goals.

EXPLANATIONS

One explanation is that statistics is mathematical and thus is harder than most other subjects. Another is that students are culturally biased against statistics. But is there an explanation involving that which is under the control of the teacher? What might explain this failure among our students based on the nature and content of the course?

Some teachers of statistics may not discern that

1. *these abstract goals are ambiguous – they may be interpreted narrowly or broadly*
2. *these goals should be interpreted broadly*
3. *teaching which achieves these narrow goals may not achieve these broader goals.*

This intellectual failure to discern ambiguity allows some teachers to proclaim allegiance to broader goals while permitting then to pursue a far narrower set of goals. This on-going ambiguity explains – in part – why students do not appreciate statistics and why other departments want statistics to be taught by their own faculty.

II. SOURCES OF AMBIGUITY IN STATISTICS

Several of the terms in these goals are ambiguous and may be taken either narrowly or broadly. When taken broadly, they are very appropriate. When taken narrowly, they miss the mark.

This systematic ambiguity is the miasma that is obscuring our vision; this wide-spread ambiguity is the swamp that is undermining our best efforts at statistical reform; this all-pervasive ambiguity is the reef upon which our hopes for improvement are being dashed.

The following terms are ambiguous. In each case, the cause of the ambiguity is indicated.

1. Estimate and judgment: Speaking mathematically, we mean 'mathematical induction' (narrow) – not pure induction (broad).
2. Inference and quantitative literacy: Speaking mathematically, we mean deductive reasoning (narrow) – not the combination of deductive and inductive reasoning (broad).
3. Explanation and factor: Speaking mathematically, we mean a statistically significant correlation (narrow) – not causality (broad).
4. Variability and quantitative literacy: Speaking mathematically, we mean the study of pure chance (narrow) – not the study of chance, influence and bias (broad).
5. Probability and chance: Speaking mathematically, we mean relative frequencies (narrow) – not the strength of a belief or the credibility of a conclusion (broad).
6. Hypothesis testing: Speaking as classical statisticians, we mean a classical test of significance (narrow) – not a Bayesian test of the null hypothesis (broad).

When combined, these sources of ambiguity make it extremely difficult for those who take these terms broadly to communicate with those who take these terms narrowly. It is no wonder that meaningful dialog is difficult.

ESTIMATE AND JUDGMENT

Students may presume that statistics will help them quantity the credibility of their estimates and judgments; students may expect measures of credibility involving pure induction. Instead, classical statistics quantifies the "confidence" they can have in their methods using 'mathematical induction'. Thus, statisticians translate credibility (broad) into relative frequency (narrow) and translate induction (broad) into 'mathematical induction' (narrow).

INFERENCE AND QUANTITATIVE LITERACY

Students may presume that statistics will help them make sounder generalizations and form stronger arguments involving quantitative matters; students may presume that statistics will help them reason inductively about the quantitative aspects of the world. Instead, statistics focuses only on valid deductive arguments whose conclusions must be true given the truth of the premises; statistics only helps students reason deductively. Students expected that inference would include induction; statisticians translate inference (broad) into deduction. (narrow)

EXPLANATION AND FACTOR

Students may expect that if education explains 67% of the variability in income that this shows that education is a causal explanation. Students may expect that if education is a significant factor in predicting salary, then this means that education is a causal factor. Statisticians are very careful to say this usage of 'explanation' and of 'factor' are technical uses which identify some feature of the correlation but which assert nothing about the causality involved. [See Milo Schield (1995) *Correlation, Causation and the Coefficient of Determination in Introductory Statistics*] But without giving students guidance as to how to move from correlation to causation, students are left without guidance or principle. Thus, statisticians translate causal explanation (broad) into correlative explanation (narrow) and translate causal factor (broad) into correlative factor (narrow).

VARIABILITY & QUANTITATIVE LITERACY
Students may presume that variability means all variation in a set of data. Instead, statistics focuses primarily on the random variation due to pure chance. Students may presume that quantitative literacy means to be literate about claims and arguments involving quantitative data. To most statisticians, quantitative literacy means understanding the role of chance in summarizing, modeling data and testing data. Statisticians translate variability (broad) into chance (narrow) and translate quantitative literacy (broad) into probabilistic literacy (narrow).

PROBABILITY AND CHANCE
Students may think of probability and chance as indicating how likely an event is – even if the event involved is a one-time, unrepeatable event. Students may intuitively think in terms of probability as a measure of the credibility of a claim – even if the claim involves a state of nature. Instead, statisticians view probability and chance as having a relative-frequency interpretation in the classical approach. Thus terms such as confidence level and level of significance should not be interpreted as measures of the credibility or error for a particular claim. Thus statisticians translate credibility of belief (broad) into an expected relative frequency (narrow).

III. CLASSIFICATION OF ELEMENTS

Among the sciences of methods, there are those that study human reasoning. These branches of epistemology include mathematics, statistics, logic and critical thinking. The following table illustrates the relation between these four sciences of human thought classified by method and by subject.

Table 1: Sciences of method about human reasoning

	Deductive Only	Comprehensive
Verbal	Logic	Critical Thinking
Quantitative	Mathematics	Statistics

Philosophically, statistics is critical thinking about quantitative claims.

Gudmund Iversen shares this broader emphasis: "The goal of applied statistics is to help students to form, and think critically about, arguments involving statistics. This construction places statistics further from mathematics and nearer the philosophy of science, critical thinking, practical reasoning and applied epistemology." [Gudmund Iversen in *Two Kinds of Introductory Courses* in *Heeding the Call for Change.* p. 29.]

David Moore shares this broader emphasis: "the higher goal of teaching statistics is to build the ability of students to deal intelligently with variation and data." [David Moore (1992) in *What is Statistics*? page 16]

Normally students classify these four method disciplines by rows – by the content of the discipline and by the department in which it is taught. But, the classification by column – by the kind of argument – is at least as fundamental.

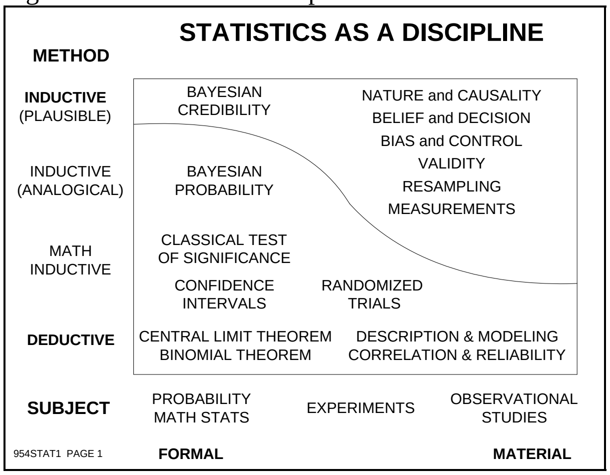
Critical thinking and applied statistics include both deductive and inductive reasoning. Both disciplines are currently in flux because they contain an inductive element. Given the current state of philosophy, the proper status of inductive reasoning is somewhat unclear. Thus, the proper status of these two disciplines is also uncertain. As a result, there is continuing pressure to eliminate the inductive aspects and to teach only what is deductively certain.

Perhaps the clearest sign of inductive reasoning in statistics is the mention of causality. A text that does not discuss this topic is unlikely to be a text that focuses on inductive reasoning. Causality is necessarily inductive; there is no valid test that conclusively proves causality – especially in an observational study. But that does not mean that all arguments about nature and causality are equally strong. Statistics must help students discern stronger arguments from weaker ones. This applies to all evaluative inferences such as the accuracy of measurements, the quality of a sample, the applicability of a model, the quality of an inference, and the strength of a conclusion about causality. Students are interested in prediction and explanation which means they are interested in causality. Without causality, statistics lacks life and purpose. [See Milo Schield (1995)]

When we think of statistics, most of us think of the subject as having several basic parts: descriptive, probability, inference and possibly modeling. But the essence of statistics is not readily determined by knowing these parts. As a discipline, statistics is described by both its methods and its subjects.

Figure 1 illustrates the various elements of statistics classified by method and subject. Experiments are most common in the physical sciences while observational studies are most common in the social sciences. However there are exceptions in both fields.

Figure 1: Statistics as a discipline



Topics below the line are generally accepted, commonly included in texts and commonly taught. Topics above the line are either not generally accepted or else are given very little emphasis in a typical introductory course.

This distinction is more than just a difference between theory and application. It is a difference in method (deductive vs. inductive) and a difference in subject matter (formal probability theory vs. the material aspects of real data found in observational studies).

Figure 2: Critical Thinking as a Discipline

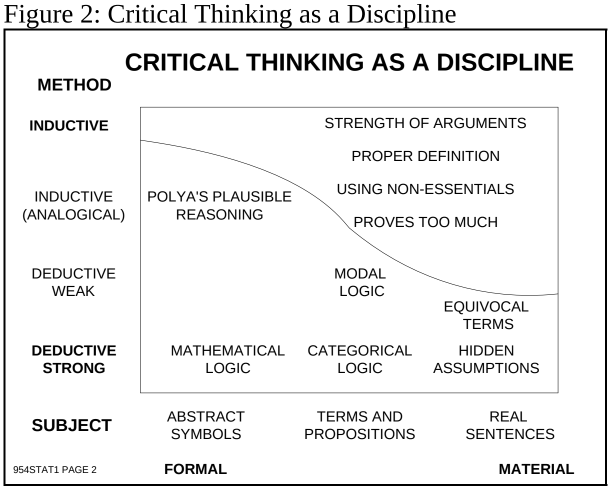


Figure 2 illustrates some of the elements of critical thinking classified by method and by subject. Critical thinking is to logic as statistics is to mathematics. Both include a deductive component, but both include an inductive component. This inductive component is what is essential to both statistics and to critical thinking.

Topics below the line are generally accepted and traditionally taught in a course on logic. Topics above the line are either not generally accepted or else are given little emphasis in a typical course in introductory logic. Those topics that are inductive (such as the formation of proper definitions) lack deductive certainty and thus are mentioned in passing but have the status of an art rather than that of a science.

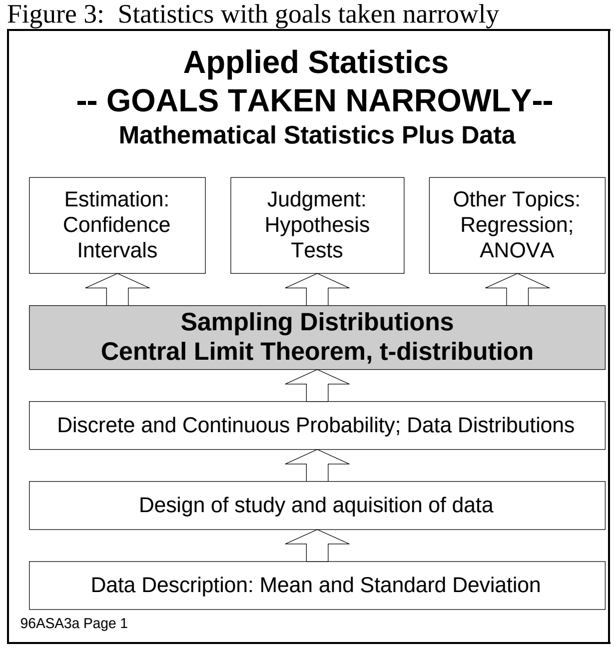
IV. GOALS TAKEN NARROWLY VS. BROADLY

It is helpful to compare applied statistics as taught with goals taken narrowly and as taken broadly.

Remember, teaching with these narrow goals is not wrong – all such topics could be appropriate in a course taught broadly. But teaching broadly – which includes the topics in any narrow approach – is required to achieve our goals.

The remaining figures are organized from bottom to top. Higher level topics depend on – and are taught after – lower level topics.

Figure 3: Statistics with goals taken narrowly



The gray background and larger box for the sampling distribution such as the central limit theorem and the t-distribution indicate their strategic importance. All subsequent topics depend on students understanding the meaning and implications of these statistical concepts. All prior topics are chosen for their ability to help students understand and appreciate the origin and truth of these key concepts. This is why probability remains such a central part in our current teaching of applied statistics.

Figure 4: Applied Statistics with goals taken broadly

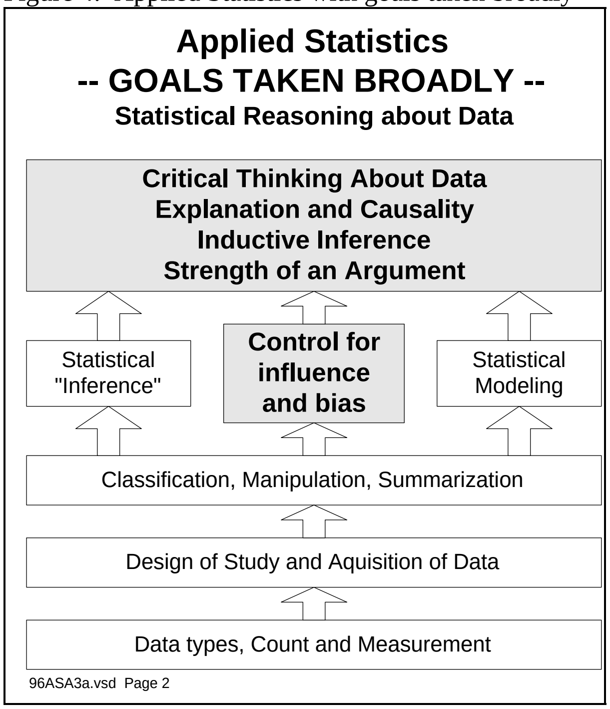


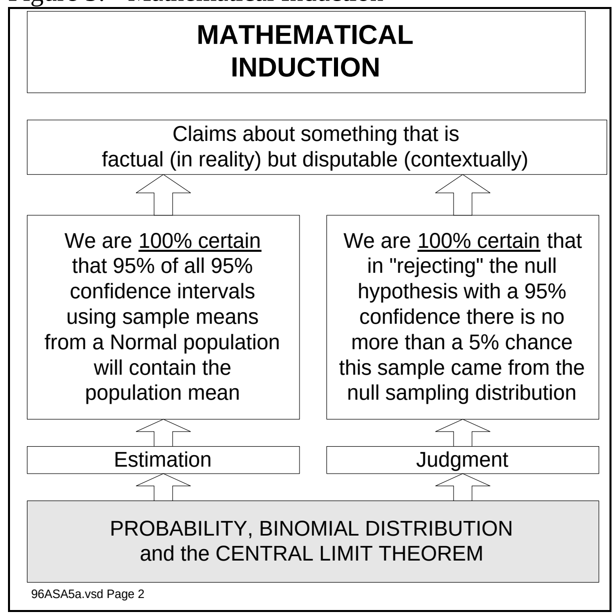
Figure 4 presents applied statistics when the goals are taken broadly – when the terms are taken at their broader generally-understood meanings rather than at their technical or translated meanings.

Again, the gray boxes indicate the items that are of strategic importance. Note that statistical inference is one of the three pillars that support our evaluation of arguments. But statistical inference is not the sole pillar nor is it necessarily the most important. The quality of our modeling and the extent to which we control for influence and bias are at least as important as the assessment of pure chance – even in small samples.

Note that the final products are not just outcomes but are actually the goals of the entire process. In this case, these goals actually determine the means. In Figure 3, the final products are not so much goals as they are outcomes. (Determining statistical independence is not so much a general goal as it is an illustration of a unique test of significance.) The final topics are selected – in large part – because they illustrate the various ramifications of various sampling distributions such as that identified by the central limit theorem.

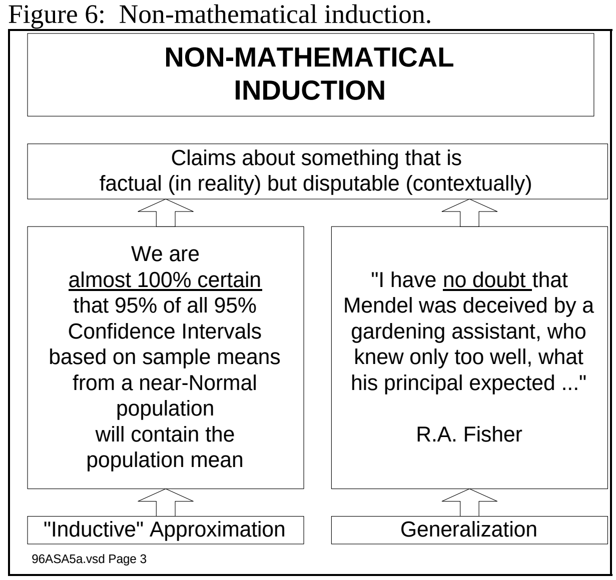
Remember, teaching with goals taken narrowly is not wrong, but it is inadequate. Consider this same issue of narrow vs. broad in the relation between 'mathematical induction' and non-mathematical induction.

Figure 5: 'Mathematical Induction'



In Figure 5, the grayed area indicates the topics that are of strategic importance. These topics give 'mathematical induction' 100% certainty about the outcome – described probabilistically. Thus, 'mathematical induction' is really deductive.

Figure 6: Non-mathematical induction.



In Figure 6, two types of induction are illustrated. One is an approximation to a deductive theorem (as commonly taught in statistics). The other involves a generalization, prediction or explanation that extends beyond the data. The data may be of crucial importance in determining the credibility of the claim – but the data are certainly not sufficient.

The well-known statement by Fisher uses “no doubt” in a way that does not involve a relative frequency interpretation. Yet, this particular statement was based on a great deal of statistical data, a particular statistical test and some specific knowledge of the person involved.

V. TOPICS MISSING OR UNDEREMPHASIZED

The emphasis on deduction and ‘mathematical induction’ determine in large part which topics will be included and the extent to which they are emphasized. The topics that are missing – or de-emphasized – say a great deal about the importance that authors and teachers place on those that are included.

(1) A topic this often under-emphasized is the control for bias. According to John Bailar (1994) in *A Larger Perspective*,

> “Our focus must not be limited to understanding the implications of pure chance (randomness). It must certainly include an understanding of bias and the methods of controlling for non-systematic error.”

(2) A topic that is often missing is resampling. Resampling has been touted for years, yet the silence on this subject is all but deafening. One wonders whether the issue is its teachability, its theoretical status or its implications for teaching statistics using a non-mathematical model.

(3) A topic that is missing is Bayesian reasoning. David Moore (1992) has given an excellent argument for why this topic should not be stressed.

> “There are, I think, good reasons not to stress Bayesian methods in beginning instruction about inference. First they require a firm grasp of conditional probability…. This [distinction between classical and Bayesian reasoning] is fatally subtle. In addition, although the subjective interpretation of probability is quite natural, it diverts attention from randomness and chance as observed phenomena in the world whose patterns can be described mathematically. An understanding of the behavior of random phenomena is an important goal of teaching about data and chance.; probability understood as personal degree of belief is at best irrelevant to achieving this goal. The line from data analysis through randomized designed for data production to inference is clearer when classical inference is the goal.”

An opposing viewpoint has been given by Berger (1980) in *Statistical Decision Theory and Bayesian Analysis*

> “most such users (and probably the overwhelming majority) interpret classical measures in the direct probabilistic [Bayesian] sense. (Indeed the only way we have had even moderate success, in teaching elementary statistics students that an error probability is not a probability of a hypothesis, is to teach enough Bayesian analysis to be able to demonstrate the difference with examples.)”

This opposing viewpoint is supported by Gudmund Iversen in *Bayesian Statistical Inference.*

A more direct reason for teaching Bayesian thinking is that Bayesian probabilities are more closely related to how strongly we believe that something is true or right than are the classical probabilities. By omitting Bayesian reasoning, we reduce our search for knowledge to only that which has deductive certainty. We seem to be saying to students “Ignore what is important but disputable; focus on what is certain. Real knowledge involves certainty.”

VI. SUMMARY

Every college graduate must be familiar with critical thinking; every college graduate should be familiar with applied statistics – when taught as critical thinking applied to quantitative data.

The ultimate question is “What do we want to teach in applied statistics?”. Do we want to teach mathematical purity, the knowledge that our claims are certain against error? Or do we want to teach reality-based relevance: the knowledge that we are willing to deal with the fact that our knowledge is contextual, that we are not omniscient, and that our inductive inferences may not always be correct.

Re-engineering statistics will be difficult. It is much easier to teach statistics as applied mathematics with right/wrong answers than it is to teach statistics as critical thinking about data where arguments are either weaker or stronger. Do we want to hear our students say “That course in statistics was really valuable. I can see how I will use statistical reasoning to think about quantitative arguments in the future.”? If so, then re-engineering statistics is what must be done. This goal is achievable and the choice is ours.

REFERENCES:

Bailar, John. *A Larger Perspective*. The American Statistician. Vol. 49, No. 1, February,1994. p.10.

Gordon, Florence and Sheldon, Editors. *Statistics for the Twenty-First Century.* The Mathematical Association of America. MAA Notes, Number 26, 1992.

Kelley, David (1994). *The Art of Reasoning*. 2nd Ed.

Hoaglin, David and Moore, David Editors (1992). *Perspective on Contemporary Statistics.* The Mathematical Association of America. MAA Notes 21.

Iverson, Gudmund (1984). *Bayesian Statistical Inference.* Sage Publications Series: Quantitative Applications in the Social Sciences.

Polya, G. (1968). *Mathematics and Plausible Reasoning*. Princeton Paperbacks. Vol. I. Induction and Analogy in Mathematics. Vol. II. Patterns of Plausible Inference. ISBN 0-691-02510.X

Schield, Milo (1994). *Random Sampling versus Representative Samples.* ASA 1994 Proceedings of the Section on Statistical Education, p. 107-110

Schield, Milo (1995). *Correlation, Causation and the Coefficient of Determination in Introductory Statistics* ASA 1995 Proceedings of the Section on Statistical Education, p. 189-194

Schield, Milo (1996). *Teaching Classical Hypothesis Testing Comprehensively.* ASA 1996 Proceedings of the Section on Statistical Education

Snee, Ronald. *What's Missing in Statistical Education?* The American Statistician Vol. 47, Number 2, May 1993.

Steen, Lynn (1992). Editor of *Heeding the Call for Change: Suggestions for Curricular Action.* The Mathematical Association of America. MAA Notes 22.

Tucker, John. (1994). *Modern Interdisciplinary University Statistics Education.* Committee on Applied and Theoretical Statistics. National Research Council. National Academy Press. Washington D.C.

ACKNOWLEDGMENTS:

Gerald Kaminski, Thomas V.V. Burnham, and Linda Schield made cogent comments on earlier drafts.

My wife, Linda Schield, edited earlier drafts for clarity, comprehension and understanding.

This paper was stimulated by Robert Hogg's assertion that "no one [really] knows the purpose of the first applied course in statistics". [ASA 1995 Orlando, FL]

This paper is based on my reflections from teaching Critical Thinking for the past 5 years and on the continuing exhortation of my close colleague Professor John Cerrito to continually focus on "helping students to read and interpret data".

Dr. Schield may be reached at schield@augsburg.edu.