STATISTICS EDUCATION: STEADFAST OR STUBBORN

Milo Schield, W. M. Keck Statistical Literacy Project. Minneapolis, MN.

Abstract: This paper claims (1) that the introductory statistics course is essentially the same in content as it was 50 years ago, (2) that statistical education has ignored many – if not most – of the content changes proposed by leaders in statistical education, and (3) that the introductory statistics course is essentially a math-stat (research-methods) course. One explanation is that statistics education is being steadfast in upholding random variation as the central idea and ultimate goal of the introductory course. Examples of how context (confounding, assumptions and bias) can influence statistical significance are shown using confidence intervals as a test of statistical significance. Statistics educators seem unwilling to adopt this into introductory statistics. An alternate explanation is that statistics educators want the introductory course to be maintained as a research methods course dedicated to upholding the purity and rigor of deductive mathematics. If so, then any attempt to introduce context may require that statistical educators actively support a separate course in statistical literacy.

Keywords Statistical literacy, context.

Introduction

Consider three claims:

1. The introductory statistics course is essentially the same – in content – as it was 50 years ago.
2. Statistical education has ignored many – if not most – of the content changes proposed by leaders in statistical education.
3. The introductory statistics course is essentially a math-stat (research methods) course.

After giving support for these claims, consider these two questions:
If these claims are true,

1. Why hasn't statistics education adopted changes recommended by its leaders?
2. What does this lack of change mean for the future of statistical education?

1. Statistics education hasn't changed

Claim #1: Statistics education hasn't changed essentially – in content – in the past 50 years. As evidence, consider the table of contents for two older statistics textbooks.

1958: *Statistics in Psychology* (5th) Garrett and Woodworth:

> PART I: DESCRIPTIVE STATISTICS (1) The Frequency Distribution; (2) Measures of Central Tendency; (3) Measures of Variability; (4) Cumulative Distributions, Graphic Methods and Percentiles; (5) The Normal Distribution; Meaning and Importance of; and (6) Linear Correlation
>
> PART II: INFERENCE AND PREDICTION: (7) Regression and Prediction; (8) The Significance of the Mean and of Other Statistics; (9) The Significance of Difference between Means & other stats; and (10) Testing Experimental Hypotheses

1966: *Elementary Statistics*, 2nd ed. Paul Hoel:

1 Nature of Statistical Methods	5 Sampling
2 Description of Sample Data	6 Estimation
3 Probability	7 Testing Hypotheses
4 Frequency Distributions	8 Correlation
	9 Regression

Most statistical educators today should be quite comfortable teaching from either of these textbooks. This lack of change provides support for the first claim that statistical education hasn't changed fundamentally in content in the past 50 years.

2. Statistics education has ignored calls for change

Claim #2: Statistics education has ignored content-change proposals made by leading statistical educators. Consider the following calls for change that have been ignored:

- Wallis and Roberts (1956): "Avoidance of mathematics is not … merely a necessity in introductory statistics; it is, we feel, a real virtue. Elementary statistics courses that draw freely on … first year college mathematics unavoidably teach mathematics at the expense of statistics, or sometimes fail to teach either The great ideas of statistics are lost in sea of algebra" (p. viii). "far reaching as have been the consequences of the t distribution for technical statistics, in elementary applications it does not differ enough from the normal distribution and does not introduce enough of a new principle, to justify giving beginners this added complexity in lieu of some other topic." (p. x).
- Selvin (1970): Disallow hypothesis tests in observational studies.
- Ehrenberg (1976): "Statistics courses are largely irrelevant: not just boring or technically difficult, but irrelevant."
- Haack (1979): Only show the test for proportions.
- Cryer and Miller (1992): Drop hypothesis tests entirely.
- Harlow, Mulaik & Steiger (1997): Show effect sizes

Consider a call for change by David Moore (1997): a leader in statistical education.

Eight Big Ideas [re-arranged] and a definition:

1. Data beat anecdotes;	6. Observation versus experiment
2. Association is not causation	7. Beware the lurking variable [confounding]
3. Importance of study design	8. Is this the right question? [definitions]
4. Omnipresence of variation	Statistical literacy: what every educated person should know.
5. Conclusions are uncertain	

Statistical educators are united in upholding the first five as big ideas that should be taught. They are conspicuously silent on upholding the last three; they are silent on whether statistical literacy should be defined by the needs of a particular group of people.

As further evidence, note that Moore's own textbooks have few instances of idea #8: "Is this the right question?" Only one instance of this was found in Moore and Notz (2009 p. 460) where the text asks this question:

> "Who is a smoker? When estimating a proportion p, be sure you know what counts as a "success." The news says that 20% of adolescents smoke. Shocking. It turns out that this is the percentage who smoked at least once in the past month. If we say that a smoker is someone who smoked in at least 20 of the past 30 days and smoked at least half a pack on those days, fewer than 4% of adolescents qualify."

In her response to David Moore's call for change, Anne Hawkins, then President of the RSS Centre for Statistical Education, issued her own call for change (Hawkins 1997a):

- I would argue in favour of 'Statistical Literacy for All', that emphasises understanding over facts and tools..." "Statistics for All, in the absence of literacy, is worthless." "Statistical Literacy for All must be the bread on which some may spread butter, jam, or even caviar.
- Sadly, some of our current practices do not suggest that we really want to do this [statistical literacy for all], nor do they always present the synergy of skills, knowledge and understanding that represent the real and adaptable natures of statistics.

In a paper, "the first R for Reasoning, Hawkins (1997b) argued

- "Let us assume that what we want to achieve is "statistical literacy" for all, and "statistical literacy plus" for some." "If there is no guarantee that more and more "plus" necessarily turns into statistical literacy, it is time for us to stop hitting our heads against a brick wall, and to engage in more radical rethinking about our approach to statistical education. Statistical education has evolved to where it is at present, but there is a case for saying that this is not the right starting point for where it should be going in the future." "The really big research question that faces us, though, is how to produce statistically literate citizens."

When a past-President of the RSS Centre for Statistical Education and a past-President of the American Statistical Association (the author of the most widely-read introductory-statistics textbook) are united in calling for more emphasis on statistical literacy the continuing lack of response by statistical educators calls for an explanation.

Wainer (2002) introduced a new graphical technique for taking into account the influence of a binary confounder. This technique was developed by Arjun Tan in 1980. See Tan (2012), Wainer's presentation was anticipated by Lesser (2001) and promoted by Schield (2004a, b, c and 2006). This is arguably the biggest breakthrough in statistical education since the introduction of personal computers and statistical calculators for it allowed teachers to demonstrate the influence of a confounder on a statistic, on a statistical association and on statistical significance – with minimal mathematics. Previously students had to complete a second course and deal with the many assumptions underlying least-squares regression before they could be shown how taking into account a related factor could influence a statistic. By his choice of title, "Making Simpson's Paradox clear to the masses," Wainer clearly recognized the value of this graphical technique. Once again, with few exceptions (Schield and Kaplan, 2011), statistical educators were steadfast in their silence on this new tool.

Consider a recent over-arching call for change by Joel Best (2001, 2002). Statistical educators should teach that:

1. Every statistic is socially constructed in the most operational sense of that term.
2. The social construction of statistics does not imply malevolence, negligence or even opportunism.
3. The social construction of statistics goes beyond chance, bias and confounding.

4. Seeing all statistics as socially constructed is essential for statistical literacy.
Note: Some of Best's concerns parallel those of Moore's "Is this the right question?"

The ASA (2012) GAISE college report mentioned statistical literacy, but was silent on whether it included any of the changes presented previously.

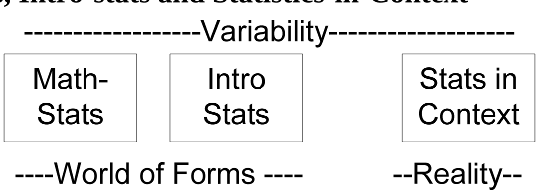
3. Introductory statistics is typically a math-stat course

Claim #3: Introductory statistics is essentially a math-stat (research methods) course. Two arguments are advanced to support this claim.

One argument notes the slavish adherence to mathematical rigor and exactitude by statistical educators. Is this mathematical need for exactitude one reason why statistical educators have not followed the advice of Wallis and Robert (1956) to stop teaching the t-test? Is this one reason why most (~70%) statistical educators don't teach statistical significance using confidence intervals? See the survey in Appendix A. Do statistical educators view a concept-based-course as being "stat-lite": something beneath the stature of a real statistician? Is this why they avoid supporting the use of a simple graphical technique to show how to take into account the influence of a binary confounder on the relationship between a binary predictor and a binary or quantitative outcome?

A second argument involves questioning the most essential idea in introductory statistics. Let's start with mathematics and math-stats. Mathematics studies quantitative patterns and structure. Mathematical statistics studies quantitative variability (with a strong focus on random variability). Figure 1 illustrates how introductory statistics might be viewed as being between a math-stats course and a statistics-in-context course.

Figure 1: Math-Stats, Intro-stats and Statistics-in-Context



Statistics should study variability in context. Statistics (the study of variability in context) is as least as far from Math-stats as math-stats is from pure mathematics. Many may disagree noting that introductory statistics is typically algebra-based with minimal focus on proofs while math-stats is typically calculus based with a stronger focus on proofs. They note that introductory statistics includes:

- how outliers can influence the mean much more than the median,
- how quantitative data may be transformed to yield a more desirable property,
- how the choice of the mean versus the median influences our summary of data,
- how error or bias can influence the data obtained, and
- how the choice of what to include in a regression influences the model.

In reply, the last item (regression) is often ignored in introductory statistics. If regression is included in introductory statistics, it usually involves just two variables whereas three are needed to demonstrate the influence of context.

Even if one accepts that introductory statistics typically covers all these topics (except the last), this doesn't justify the claim that introductory statistics focuses adequately on variability in context so that introductory statistics is further from math stats than from statistics in context. What is the gap between introductory statistics and statistics in context?

Consider this single question: Have you EVER seen an introductory statistics text book that showed how statistical significance could be influenced by what was – and was not – taken into account, how statistical significance could be influenced by how things were counted or measured, and how statistical significance could be influenced by bias?

If you have never seen any introductory statistics textbook that showed how these sources of variability can influence statistical significance, then it seems that the conclusion stands: introductory statistics (research methods) is closer to math-stats than to statistics in context.

Students completing an introductory statistics course are never shown how any of these influences on a statistic affects statistical significance. If students leave introductory statistics thinking that the presence or lack of statistical significance is a "fact" much like whether an integer is even or odd, are they in error? They have never been shown a situation in which statistical significance was influenced by anything aside from the sample size and variability in random sampling or random assignment.

To repeat: introductory statistics courses ignore "Context": the fact that statistics are different from numbers. Statistics are based on something in reality; numbers are Platonic abstractions. In broadest terms, 'Context' includes all the influences on a statistic. These can be grouped into three categories:

- context: what is taken into account (controlled for) and what is not (confounding)
- assembly/assumptions: how groups are defined and how quantities are measured
- error/bias: how a statistic can be influenced by a subject's motives (subject bias), by the means or criteria of measuring or counting (measurement bias) and by the way in which a population of interest was sampled (sampling bias).

Note that 'context' is being used in two ways: as the overall group (all influences on the size of a statistic) and as a member of that group (those influences that were taken into account and those that could have been taken into account but were not). This two-level usage of 'context' is similar to how we use "man": 'Man' is used to indicate human (men and women) and to indicate the male within the group of humans.

4. Statistics education: Steadfast or Stubborn

So is statistical education being steadfast or stubborn in resisting calls for change? It certainly depends on what is essential to introductory statistics. What are the most important topics in introductory statistics? McKenzie (2004) surveyed statistical educators. The following data involves two questions:

1. What are the core concepts in introductory statistics? See column 2
2. What are the top three core concepts in introductory statistics? See column 3.

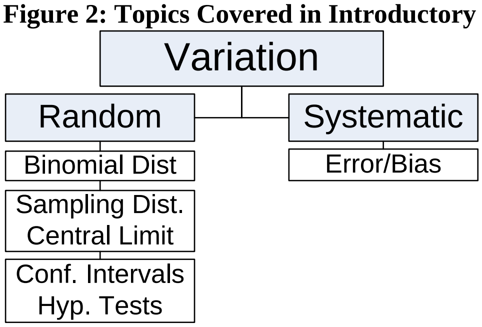
Here is data for the most common answers shown as a Percentage of all respondents. Obviously multiple answers are permitted for both questions.

Table 1: Core Concepts in Introductory Statistics

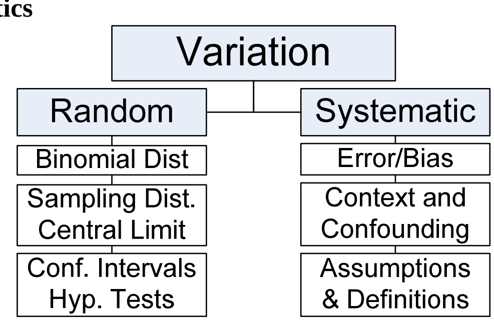
Topics/Concepts	Core Concept	Top 3 Importance
Variability	96%	75%
Association-causation	82%	31%
Randomness & statistical significance	77%	14%
Data, experiment-observational study	75%	14%
Sampling distribution	71%	24%
Hypothesis tests: Critical value, p-value, power	64%	25%
Confidence Interval	63%	22%
Random sample	63%	12%

In both cases, variability (variation) was the highest-ranking core concept. Figure 2 illustrates the topics covered in introductory statistics now – and how they would need to be if context were fully included.

Figure 2: Topics Covered in Introductory Statistics



Introductory Statistics: Now



Introductory Statistics: How it would be

If introductory statistics were to focus on variation in context, it would look more like the right side of Figure 2. This may seem like a small change and one that statistical educators might readily embrace.

5. The most important idea in introductory statistics (Part 1)

Hypothesis #1: Variation is not the most important concept in statistical education – but it is the genus of that concept. **Random variation** is the core idea. This may explain why statistical educators may have responded as they did to the above survey. They may not have seen how the study of confounding could be tied in with statistical significance. This may also explain why statistical education has rejected so many calls for change. Statistical education may appear stubborn, but perhaps it is being steadfast in upholding random variation as the central idea and ultimate goal of the introductory course.

As another piece of evidence, note that 'confounding' was not even an option in McKenzie's survey. Yet Tintle et al (2013) claimed that "confounding and variation are **the two** substantial hindrances to drawing conclusions from data – and the two major themes of statistical analysis." How can 'confounding' be absent on McKenzie's survey of key concepts and yet be named as one of the two major themes of statistical analysis? This ambivalence reflects a very deep underlying tension in statistical education.

If hypothesis #1 is true, then to be successful in changing introductory statistics, any proposed change should uphold **random variation** and **statistical significance** as the core ideas – the crown jewels – of statistics education.

6. Adding Context to Introductory Statistics

To present statistics as *numbers in context*, the introductory statistics course would have to be enhanced with a greater focus on a big idea: context may influence the size of a statistic, the size of a statistical association, or the statistical-significance of an association. To repeat, these additional sources of influence can be classified into three groups:

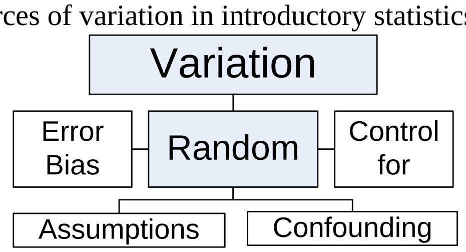
- context (what is controlled for; confounding),
- assumptions or assembly (definitions of groups and measures) and
- error or bias

Adding systematic sources of variation to introductory statistics may have big benefits:

- improve student retention of key ideas,
- improve student attitudes on the value of studying statistics,
- uphold context – not variability – as the essential difference between statistics and mathematics.

Figure 3 shows how topics could be organized so statistical education could maintain random variation as the central idea.

Figure 3: Presenting sources of variation in introductory statistics.



Introductory statistics could be extended to study all sources of variation: assumptions, confounding and bias. In each case, introductory statistics should show how each source of variation can influence statistical significance. Here context is broken into two parts: those items already controlled for by comparisons, study design and models, and those related factors that were not controlled for (confounders).

There are two distinct ways of introducing context into an introductory course. Both ways would start by taking into account the influence of related factors: confounders, changes in definitions or measures, and the influence of various kinds of bias. They two approaches differ in how they test to see if the resulting change or difference is statistically-significant.

1. One way is to simply use the statistical tests presented in introductory statistics to see if the difference is statistically significant.
2. The second embraces the use of confidence intervals to determine statistical significance. Non-overlapping 95% confidence intervals would be sufficient for statistical significance.

The second approach (overlapping confidence intervals) is problematic for two reasons.

1. The overlap misconception: Kalinowski (2013) noted that “38% of medical researchers, 31% of behavioral neuroscientists and 25% of research psychologists believed that a statistically significant result corresponds to the 95% confidence

intervals just touching. When the [95% confidence] intervals are just touching, the corresponding p value is actually .01."
2. The use of confidence intervals obscures the importance of power.

Once again there is a tension between mathematical rigor and quick comprehension.

Schield (2004a, b, c and 2006) shows how the second method can be used to demonstrate the impact of a binary confounder on statistical significance using confidence intervals. Appendices B and C show how assembly (assumptions) and bias can influence statistical significance using confidence intervals. .

7. Concerns

Pearl et al (2012) noted that "As the field and practice of statistics has changed, it has become more difficult to provide an agreed upon list of specific topics or procedures that all students should learn." This is a most important development in statistical education. It may reflect a willingness to rise above a math-stats research-methods approach to introductory statistics. It may also reflect an unwillingness to abandon the security of an introductory course that is grounded in rigorous and challenging mathematics with deductive certainty.

Schield (2004b) spoke to an invited group of international statistical educators hosted by the IASE in Lund, Sweden. Schield presented a graphical technique showing the influence of a binary confounder on the association between a binary predictor and a ratio outcome. He also used this graphical technique – along with the associated confidence intervals – to show how controlling for a confounder could influence statistical significance. After his presentation, he surveyed the audience on two questions:

Q1. Should students be shown that statistical significance can be influenced by a confounder?
Responses: Strongly agree (8), agree (7), neutral/indifferent (1), disagree (0).

Q2. Should introductory statistics teach students more about confounding even if that means less time for statistical significance?
Responses: Strongly agree (0), agree (7), neutral (4), disagree (5).

You can see the shift in answers from the same group of people. If focusing more on context means less focus on statistical significance then that was a big problem.

Schield (2013) piloted a draft of this paper with a small group of statistical educators. They were shown how statistical significance could be influenced by confounding, by bias and by assumptions or assembly. They were shown these simple ways of presenting these influences on statistical significance to students. Attendees were generally reluctant to adopt these changes into their classes. When asked why, one leader in statistical education said "When students see how easily statistical significance is influenced by other factors, this may bring our discipline into disrepute."

This unwillingness to teach students about the possible influence of confounders on statistical significance may reflect stubbornness among statistical educators. Today's leaders in statistical education (Tintle et al, 2013) have called for change in the second course with the primary focus on confounding and variation. If this call to focus on confounding in the second course is not carried over into the first course, then statistical educators may be viewed as being steadfast to a math-stats approach to introductory statistics while stubbornly resisting any effort to add relevant context that opens the door to inductive reasoning.

8. The most important idea in introductory statistics (Part 2)

Rather than argue that statistical educators are unwilling to introduce context into introductory statistics courses, consider another possibility. Most statistical educators are well-trained in mathematics. As statisticians, they know that in practice much of statistical analysis will involve incomplete data, mislabeled variables, miscoded values, models that don't quite fit, etc. These matters are usually addressed in the second course in modeling. But they are grounded in the mathematical theory underlying statistical inference.

Hypothesis #2: Perhaps statistical educators are proud to offer the introductory course as the closest to a math-stats course that can be done without calculus, without proofs and without much depth in the theory of probability. Introductory statistics may be a way that statistical educators pay homage to the power and beauty of mathematics – uninhibited by most matters of context.

In this world of forms and relationships, some differences don't really matter. Thus, whether data is obtained from an experiment or an observational study has no effect on how to model the data or on whether a given difference is statistically-significant. Using correlation as evidence for causation does not arise because causation is not a mathematically-accessible topic. We simply say association is not causation.

Certainly there is a possible benefit to showing students in quantitative majors the stark beauty, the deductive power and the certainty of higher-level mathematics. In such a course, we can focus on demonstrating the central limit theorem – instead of proving it.

The survey results reported in Appendix A support this hypothesis. Less than a fifth of statistical educators surveyed have ever showed how statistical significance could be influenced by other factors.

If this allegiance to pure mathematics were central to the design of introductory statistics, then most of the resistance to change by statistical educators is no longer a case of being stubborn – it is a case of being steadfast. Statistical educators would be justified in resisting attempts to introduce the influence of confounding, or assumptions and of bias into the introductory course since it is intended to be a research methods course.

9. Conclusion

If statistical educators want to introduce context and confounding into the first statistics course, they face a choice:

1. Back off on upholding introductory statistics as a research methods course. We already talk about subject bias and research bias: two topics that don't really belong in a math-stats course. Spend less time or showing how the central limit theorem is obtained; spend more time showing how statistical significance is influenced by related factors.
2. Support offering a separate statistical literacy course that shows how context can influence the size of a statistic, the size of an association, or the statistical significance of an association. Rossman (2007) has noted that "This [the GAISE Guidelines] is a great list of goals; These goals are hard to attain – and understand 'why'; You simply can't achieve these goals in one course if you also teach

> a long list of methods. Most students would be better served by a Stat 100 course than a Stat 101 course."

Statistical educators should decide how to include the influence of related factors – such as confounding – in an introductory course. Simply ignoring the influence of context is a disservice to our discipline: a discipline in which context distinguishes us from pure mathematics. If statistical educators want to preserve Stat 101 as a research methods course, then they may need to support a Stat 100 course in statistical literacy.

10. Acknowledgements

To Marc Isaacson, Cynthia Schield and Thomas V. V Burnham for comments and recommendations.

REFERENCES

ASA (2012). GAISE Report.

Best, J. (2001). Lies, Damned Lies and Statistics. Univ. California Press.

Best, J. (2002). People Count: The Social Construction of Statistics. Augsburg College. Copy at www.statlit.org/pdf/2002BestAugsburg.pdf

Cryer, J. and R. Miller (1992). *Statistics for Business: Data Analysis and Modelling*. (The Duxbury advanced series in statistics and decision sciences). First edition.

Ehrenberg, A. S. C. (1976). We must preach what is practised: a radical review of statistical teaching. *Journal of the Royal Statistical Society*, Series D, 25(3), 195–208.

Haack, (1979). *Statistical Literacy*. Duxbury Press.

Harlow, L., Mulaik S and Steiger J. (1997). *What If There Were No Significance Tests?* Multivariate Applications Series. Psychology Press.

Hawkins, A (1997a). The First R – for Reasoning. 1997 Annual Mathematics Teachers' Conference organised by the Mathematics Panel of the Northern Ireland Educational Support Unit. Copy at www.statlit.org/pdf/1997HawkinsRSSCSE.pdf

Hawkins, A (1997b). Response to David Moore's "New Pedagogy and New Content: The Case of Statistics", *International Statistical Review*, 199. Copy at www.stat.auckland.ac.nz/~iase/publications/isr/97.Moore.pdf

Kalinowski, Pawel (2013). Finding Statistics that Communicate: How do People Understand Confidence Intervals? PhD Thesis, School of Psychological Science, LaTrobe University. Bundoora, Victoria, Australia.

Lesser, L. (2001). "Representations of Reversal: Exploring Simpson's Paradox." In Albert A. Cuoco and Frances R. Curcio (Eds.) *The Roles of Representation in School Mathematics* Yearbook. National Council of Teachers of Mathematics, 129-145

McKenzie, Jr., John (2004). Conveying the Core Concepts. ASA Proceedings of the Section on Statistical Education. www.statlit.org/pdf/2004McKenzieASA.pdf

Moore, D. (1997). Statistical Literacy and Statistical Competence in the 21st Century [slides]. www.statlit.org/pdf/1997MooreASAslides.pdf

Moore, D. (1997). New pedagogy and new content: the case of statistics. *Int. Statistical Review*, 65,123-165. www.stat.purdue.edu/~dsmoore/articles/PedagogyContent.pdf

Moore, D. and W. Notz (2009). *Concepts and Controversies*. 7th ed. W. H. Freeman.

Pearl, D., Garfield, J., delMas, R., Groth, R., Kaplan, J. McGowan, H., and Lee, H.S. (2012). Connecting Research to Practice in a Culture of Assessment for Introductory College-level Statistics.
[Online: www.causeweb.org/research/guidelines/ResearchReport_Dec_2012.pdf]

Rossman, A. (2007). Seven Challenges for the Undergraduate Statistics Curriculum in 2007. USCOTS. Copy at www.StatLit.org/pdf/2007RossmanUSCOTS6up.pdf

Schield, M. (2004a). "Statistical Literacy and Liberal Education at Augsburg College," *AAC&U Peer Review*, 6(3). Copy at www.StatLit.org/pdf/2004SchieldAACU.pdf

Schield, M. (2004b). Statistical Literacy Curriculum Design, 2004 IASE Curriculum Roundtable, Lund Sweden. Copy at www.statlit.org/pdf/2004SchieldIASE.pdf

Schield, M. (2004c). Three Graphs to Promote Statistical Literacy. ISI Copenhagen. Copy at www.statlit.org/pdf/2004SchieldICME.pdf

Schield, M. (2006). Presenting Confounding and Standardization Graphically. *STATS*, ASA. Fall 2006. pp. 14-18. Copy at www.StatLit.org/pdf/2006SchieldSTATS.pdf.

Schield, M. (2013). Adding Context to Introductory Statistics. A talk at the Twin Cities Chapter of the ASA held at Augsburg College. Copy of slides at www.statlit.org/pdf/2013-Schield-ASA-TC6up.pdf

Schield, M and D. Kaplan (2011). Teaching Confounding and Adjustment through the Common Core Standards. USCOTS Poster. Copy at at www.statlit.org/pdf/2011-Kaplan-Schield-Poster-USCOTS.pdf

Selvin, H (1970). A Critique of Tests of Significance in Survey research in *The Significance Test Controversy*. Morrison, D.E. & Henckle, R.E. , Eds (1970). Aldine

Tan, A. (2012). How I Created a Trapezoidal Display of Simpson's Paradox. Letter to M. Schield. Copy at www.statlit.org/pdf/2012-Arjun-Tan-Simpsons-Paradox.pdf

Tintle, N., B. Chance, G. Cobb, A. Rossman, S. Roy, T. Swanson & J. VanderStoep (2013). Challenging the State of the Art in Post-Introductory Statistics: Preparation, Concepts, and Pedagogy. ISI. Copy at www.statistics.gov.hk/wsc/IPS032-P1-S.pdf

Wainer, H. (2002). "The BK-Plot: Making Simpson's Paradox Clear to the Masses." *CHANCE*, 15 (3): 60-62

Wallis, A and H. Roberts (1956). *Statistics: A New Approach*. The Free Press, Glencoe, IL

APPENDIX A: Statistical Educator Survey

The following data was obtained by a show of hands from the 50 attendees at the Statistical Literacy 2013 session at the 2013 ASA Joint Statistical Meeting.

Have you EVER taught a traditional intro stats course where you SKIPPED OVER the following?

1. the difference between experiment & observation study ~10%
2. how random assignment controls for confounders ~20%
3. the T-test: one or two population 5/50

Have you EVER taught a traditional intro stats course where you SHOWED that Statistical Significance …

5. can be tested using confidence intervals ~30%
6. can be changed by controlling for a confounder 8/50
7. can be changed by the presence of bias 7/50
8. can be changed by re-defining a group or measure 4/50

APPENDIX B: Influence of Bias on Statistical Significance

Normally bias cannot be measured – or even detected. But, suppose it could be.

Example #1: Suppose that men and women are asked about their income and men averaged $11,000 more than women. For the sample size suppose that $5,000 was the 95% margin of error. Thus the stated difference in incomes would be statistically significant.

Figure 4: Subject bias and statistical significance

	$5,000	is the 95% margin of error			
Income	Men	Women	Diff	Overlap	Stat.Sig
Stated	$62,000	$51,000	$11,000	No	Yes
Actual	$53,000	$51,000	$2,000	Yes	No

But suppose that the men in this survey were more likely to overstate their income than were women. Suppose the bottom row shows the actual results in fact. Now the $2,000 difference is no longer statistically significant.

Result: In this case, subject bias created a statistically-significant difference that was spurious.

Example #2: Suppose that among the survey responders, men and women reported a $2,000 difference in their income. If the 95% margin of error were $3,000, then the reported difference would not be statistically-significant.

Figure 5: Sampling bias and statistical significance

	$3,000	is the 95% margin of error			
Income	Men	Women	Diff	Overlap	Stat.Sig
Responders	$53,000	$51,000	$2,000	Yes	No
Population	$62,000	$55,000	$7,000	No	Yes

But suppose that those who were rich were much less likely to respond to surveys than those who were not – and those who were rich were more likely to be men. Suppose the bottom row contained the data from the entire population. In this case, the $7,000 difference between men and women would be statistically-significant.

Result: In this case sampling bias made a statistically-significant difference appear as though it were statistically-insignificant.

APPENDIX C: Influence of Assembly/Assumptions on Statistical Significance

Mathematicians and statisticians are the first to agree that the choice of the assumptions and definitions can readily determine the outcome. Consider how this choice of how to assemble subjects into groups or how to measure quantities can influence statistical significance.

Example #1: Suppose that bullying is defined as the use or threat of physical force. Using this definition, boys are 30 percentage points more likely to bully than are girls. If the 95% margin of error is 5 points, then this difference is statistically-significant.

Figure 6: Defining measures and statistical significance

Middle-schoolers	5%	is the 95% margin of error			
BULLYING	Boys	Girls	Diff	Overlap	Stat.Sig
Physical only	40%	10%	30%	No	Yes
Physical & Social	45%	44%	1%	Yes	No

Now suppose that bullying is redefined to include social ostracism, name calling and teasing. Now girls are almost as likely as boys to bully and the difference in bullying prevalence between boys and girls is no longer statistically-significant.

Result: In this case changing the definitions transformed a statistically-significant difference into one that was not statistically-significant.

Example #2: Suppose that men and women are surveyed on their attitude toward fishing using three categories: dislike, neutral and like. The 95% margin of error is 6 points.

The three categories are summarized into two groups. In the first case, 'like' excludes 'neutral'. The 10 point male-female difference is not statistically-significant.

Figure 7: Defining groups and statistical significance

	6%	is the 95% margin of error			
Fishing	Dislike	Neutral	Like	% who like*	% who like**
Men	30%	30%	40%	40%	70%
Women	50%	20%	30%	30%	50%
* Excludes netural			Overlap	Yes	No
** Includes neutral		Stat. significance		No	Yes

But, in the second case, all those who did not choose 'dislike' are considered to be in 'like'. Now the 20 point male-female difference is statistically-significant.

Result: In this case, changing the process by how groups are constructed from sub-groups transformed a statistically-insignificant difference into one that is statistically-significant.