Analyzing Numbers in the News: A Structured Critical-Thinking Approach

Milo Schield, Augsburg College
W. M. Keck Statistical Literacy Project

Abstract

A quantitatively literate reader should be able to understand and evaluate numbers in the news. Some news-based courses focused on unstructured critiques in order to avoid structured formulaic approaches. Steen noted that these idea courses made them "more difficult for teachers to teach and for students to master." An alternate approach focuses on questions or topics that provide a general structure without degenerating into mindless formalism. This paper presents an approach used in teaching a statistical literacy course by the W. M. Keck Statistical Literacy Project. This approach focuses on 10 distinct elements found within most news stories that use statistics as evidence for non-statistical conclusions. This approach is evaluated in two ways: by comparison with similar approaches and by the ability to distinguish different levels of analysis such as quantitative reasoning, statistical thinking, quantitative literacy and statistical literacy.

1. News-Based Numeracy Courses

Quantitative courses may focus on doing (design, execution, analysis and summarization of surveys, experiments or studies) or on the interpretation and evaluation of the numbers in reports or news summaries of these surveys, experiments or studies. This paper focuses on the latter. These quantitative courses will be generally described as 'numeracy', 'statistical literacy' or 'quantitative literacy' (QL) courses.

Teaching a quantitative course based on evaluating numbers in the news has been recommended by educators in statistics. The American Statistical Association's GAISE college report (2006) suggests assessing statistical literacy by students "interpreting or critiquing articles in the news and graphs in media." Steen (2004, p 47) noted that "The essence of QL is to use mathematical and logical thinking in context."

Numeracy courses that use quantitative news stories can be classified as news-enhanced or news-based. Both types may carefully select news stories to illustrate particular ideas or techniques. News-based numeracy courses also expect students to handle any news story involving numbers as part of the course while news-enhanced courses do not. This paper deals with news-based numeracy courses.

News-based numeracy courses can be classified as formulaic and non-formulaic. In formulaic courses, students are given some aids in analyzing news stories, while in non-formulaic courses they are not.

2. Non-Formulaic News-Based QR Courses

Teaching a non-formulaic news-based QR course based on numbers in the news has been tried by teachers in various venues. Here are some examples.

2.1. Detecting Statistical Doublespeak

Hack (1976) pioneered teaching statistics using a news-based course. "A new statistics course is being taught at the University of Kentucky which does not require students to learn the symbolic language of statistics (the formulas). Students who will never take a traditional "methods" course in statistics learn to become better consumers of statistic fed them daily by the different news media." See Appendix A for details.

2.2. The Chance Project

In 1992, the Chance project at Dartmouth was funded by NSF. This project, led by Laurie Snell, was based on using current news stories as the basis for presenting statistical summaries and probabilities.

Snell (1999) noted, "We designed the course in a very simple way. In a typical class, we start by having the students in groups of three or four to read an article in the current news that uses concepts of probability or statistics and answer a two or three discussion questions relating to the article."

Snell [verbal communication] concluded that the project was more difficult to teach than anticipated. It required more commitment and preparation on the part of the teachers than required for a traditional topic-based course in mathematics or statistics.

These difficulties may explain why the Chance course was extended to include "what Joan Garfield has dubbed a 'chance enhanced' course: a standard course enriched with discussion of chance news items."

2.3. Madison's Focus on News Stories

Madison (2006) has argued that a quantitative literacy class must be based on numbers in context. He argued that the class materials that provide the contexts for the mathematics and statistics problems must be authentic and the source articles must be fresh. Madison argued that the characteristics for QL-friendly courses should include "freshness, few formal algorithms, venues for continued practice, and emphasis on number sense."

Madison (2006) identified 17 conjectures involving QR/QL. Madison concluded that "template problems are antithetical to QL." See Appendix B for details.

2.4. Burnham's Unstructured Approach

At a statistical literacy workshop, Burnham (2003) provided a definition of statistical literacy (Appendix F): "The habit of noticing without specific prompting the strengths and weaknesses of such claims and reports of statistical information and arguments based thereon as commonly appear in the non-technical media." For Burnham, students should just be asked to "Comment.".

2.5. Steen's Concern

Steen (2004, p. 39) noted that "Earlier innovative, QL-type courses "had one thing in common that contributed to their remaining a small elective rather than a major requirement – they were designed specifically to focus on ideas – generally QL-like ideas – rather than techniques. This made them more difficult for teachers to teach and for students to master, and for that reason they thrived only in special niches out of the mainstream of college mathematics."

This benefit of teaching formulaic techniques may explain why the MAA approach to QL focuses on the topics to be taught. According to Gilman (2006), "*There is consensus that the mathematical skills necessary to be quantitatively literate include elementary logic, the basic mathematics of financial interest, descriptive statistics, finite probability, an elementary understanding of change, the ability to model problems with linear and exponential models, estimations and approximation, and general problem solving*."

Some may wonder if a techniques-based course is really a QL course. Some may certainly wonder if these topics are proximately related to the numbers found in everyday news stories. But this is a side issue.

The central question is this:

> *How can a news-based numeracy course maintain freshness without becoming a unique but unrepeatable niche course and without degenerating into a techniques-based algebra-like mathematics course?*

2.6. Case-based teaching

This dilemma is not unique to quantitative literacy. Any teacher that uses a case-based approach faces the same problem. They want the student to be able to comment on an individual case without expecting a formulaic answer, yet they realize that students need some structure to help them.

To help students make the transition from technique-based stories or cases to unprompted stories or cases, general templates are provided to help them organize their thinking. In analyzing human events, reporters are advised to think in terms of the five Ws: Who, What, Where, When and Why. Business cases are analyzed in terms of Strengths-Weaknesses-Opportunities-Threats (SWOT analysis). Law cases are briefed by a review of the relevant facts, issues, rules and reasoning. A more detailed briefing may include Procedural History, Legal Issue, Facts of Case, Statement of Rule, Policy, Dicta, Reasoning, Holding, Concurrence and Dissents.[1]

One wonders whether a similar set of categories or questions might help students deal effectively with a wide variety of news stories involving numbers.

3. Formulaic News-Based QR Courses

A more structured approach to teaching QL is to focus on the arguments involved in news stories. What are the questions a numerate reader would raise? What are the skills a numerate reader would use? What are the topics a numerate reader would review? By focusing on the general questions, skills or topics, one might avoid Steen's concern of being unique but unrepeatable while upholding Madison's goal of maintaining uniqueness and freshness without becoming formulaic.

The following are three lists of general questions, skills or topics to help students analyze numbers in the news.

Gall (1999) identified 10 types of questions that are relevant to analyzing statistics and chance in the news. See Appendix C.

Bracey (2006) identified 32 principles of data analysis that should be used when analyzing social science data. See Appendix D.

Lutsky (2007) identified 10 foundational QR questions that are relevant in analyzing numbers and data. See Appendix E.

4. Statistical Literacy

This paper presents a fourth approach to upholding Madison's goal of freshness while attending to Steen's concern for repeatability. This approach is a key component of the W. M. Keck statistical Literacy project at Augsburg College. It has been used in teaching a news-based statistical literacy course (GST 200). This course, taught since 1998, is an applied critical thinking course designed to help students analyze numbers in news stories. Most of the students are in non-quantitative majors – majors that do not require a specific mathematics course (majors such as English, communications, history, political science, journalism, political science, music and philosophy). Most of these students have little experience in analyzing essays that use numbers as evidence.

The goal of this course is to sensitize a student to those things that influence the strength or weakness of an argument involving numbers as evidence. These students are generally unprepared to meet the Burnham challenge: "comment." They need to be prompted.

[1] http://www.lawnerds.com/guide/briefing.html

The following analysis presents 10 categories. The approach is a spiral or hierarchical approach. Within each category or topic the general topics or questions involve more detailed subsidiary topics or questions.

The 10 specific "comment" prompts are to comment on (1) the nature and claims of the essay, (2) the causal connections, (3) the numbers used, (4) the ratios and models, (5) the study design, (6) the influence of randomness, (7) the plausible sources of error or bias, (8) the plausible confounders, (9) the opportunities for social construction, assembly, ambiguity and prevarication, and (10) the overall strength/weakness of the main argument and any associated decision.

1) Critical Thinking

a. What kinds of numbers are involved: factual data, (e.g., sports, finance, weather), a medical or educational diagnostic test, an experiment or clinical trial, the results of a model or prediction, a general survey, or an observational study or report?
b. If inferences are involved, what kinds are they: deduction, hypothetical-deductive or induction (practical reasoning)?
c. If claims are involved, what kind are they: generalization, prediction, specification, causation?
d. Does the story have a point? What is it?
e. Does the point assert causation (e.g., cause, effect)?
f. Does the story involve words that imply causation? (Does it use action words like 'cut' or 'prevent')
g. If the story uses modals (can, may or might), does the uncertainty involve an individual (someone will win; you may win), an outcome (It may rain), the cause (X may be the cause) or is it ambiguous?

2) Causation [If applicable]

b How could the association be causal?
c. What 3rd factor mechanism may be involved?
d. Could the association involve reverse causation?
e. What factors are already taken into account?

3) Statistical Association

a. What numbers or numerical association gives the strongest support?
b. Is the association qualitative (e.g., more/less, bigger/smaller) or quantitative (e.g., % more/less)?
c. Is outcome (predictor) categorical/qualitative (student or rich) or quantitative (weight or income)?
d. If the numbers involve ranks, percentiles or percentage points, what information might be missing?
e. Is the outcome actual (counted/measured) or predicted/modeled (% attributable to Y)?
f. Does the data refer to the future explicitly (prediction) or implicitly (chance/probability/likely)?
g. Does the association involve a slope (As X increases, Y increases; For every 10% increase in X, Y increases by P%) or a comparison of slopes?

4) Context: Comparisons, Ratios & Models

a. Are counts used in place of ratios?
b. If ratios are used are they appropriate? Should ratios be compared? Is there any confusion of the inverse?
c. Are measures compared for groups? If so is the effect size given?
d. Are two sets of measurements being compared? If so is the slope or correlation coefficient given?
e. Are the numbers based on a model? Are they used outside the model range? Do they imply causation?

5) Study design

a. Is the association experimental or observational? If experimental, were subjects randomly assigned?
b. Is the association longitudinal or cross-sectional? If longitudinal, are the subjects a cohort?
c. Does the data refer to past (percentage/rate) or to the future (chance/probability/likely)?
d. Is the study single or double-blind?

6) Randomness

a. Are the subjects homogeneous or heterogeneous in relation to the factors of interest?
b. Does the study involve a sample and was the sample randomly selected? If so, is the margin of error presented; is the association "statistically significant"? If not, is the sample size given?

7) Error or Bias

a. subject bias (difference due to subject awareness)
b. measurement bias (question design, researcher awareness of subjects in treatment group)
c. sampling bias (differences between the treatment and control group due to the sampling methodology)

8) Context: Confounding

Recall that a confounder is an associated factor not controlled for in the study.

a. What kinds of confounders does study design resist?
b. What are some plausible confounders?
c. Is X a plausible confounder?

Note that the last two involve hypothetical thinking .

9) Assembly in Definition and Presentation[2]

a. What opportunities for assembly involve definitions of groups, conditions or events?
b. What opportunities for assembly involve choice in type and/or basis of comparison?
c. How would a statistic or association change if a group, cutpoint or measure was redefined?

10) Critical Thinking: Evaluation and Decision

a. How readily can numerical evidence (association) be influenced by other factors?
b. Could the quality of the study have been improved?
c. How much support does the numerical evidence give to the point of the essay?

---

[2] See Best (2008a, 2008b)

5. Comparative Evaluation of Topics

In Table 1, the categories in Schield's approach are compared with those in three other formulaic approaches.

Schield's 10	Gal's 10	Lutsky's 10	Bracey's 20 (Excluding Tests)
1. Critical Thinking: Type of argument, inferences.			
2. Causation	8. Are claims sensible?		14. Correlation not always causation. 15. Correlation may be meaningless. 25. Rising test scores do not necessarily mean rising achievement.
3. Numbers, Association	4. How is data distributed? 5. Are statistics appropriate?	1. What do numbers show?	1. Do the arithmetic 2. Show me the data 5. Rhetoric and numbers must match.
4. Context: Comparisons, ratios and models		2. How representative? 3. Compared to what? 5. What is the effect size?	9. Numbers/scores vs. rates 11. Ranks vs. scores 28. Comparisons must fit data.
5. Study Design	1. Where did data come from?	7. What is study design?	
6. Randomness: chance, margin of error, statistical significance	2. Valid inferences from sample? 7. How was chance calculated?	4. Statistically significant?	13. Statistical vs. practical significance.
7. Error or bias	3. Reliable or accurate measures?	6. Source of numbers? 9. Who is in the sample?	3. Beware of selectivity in groups
8. Context: Confounding	10. Alternate explanations?	10. Controlling for what?	4. Compared groups must be similar/ 10. Groups compared over time must remain similar as years go by. 12. Watch for Simpson's paradox.
9. Assembly in definition and in presentation	6. Is graph appropriate?	8. Variable operationalized?	8. Is "average" the mean? 16. What does graph really say? 22. Pass/cut point is always arbitrary. Ensure it is not capricious.
10 Summary & Evaluation	9. Additional info needed?		6. Beware of claims that public schools are always to blame. 7. Beware of simple explanations for complex phenomena. 23. Always ask, "So what?"

Table 1: Comparison of Three Approaches with Schield

Differences in emphasis are obvious. Schield focuses more on arguments (topic 1). Bracey (2006) focuses more on tests. Gal and Lutsky focus less on the arguments and more on the data. Lutsky and Bracey provide more detail in some areas than Schield and Gal. This may indicate that Schield and Gal combine topics that should remain distinct.

6. Clarification Benefit

The Schield approach may have benefit if it clarifies distinctions between closely-related ideas such as numeracy, quantitative literacy, statistical thinking and quantitative reasoning. To see this, consider the 10 Schield topics as shown in Figure 1.

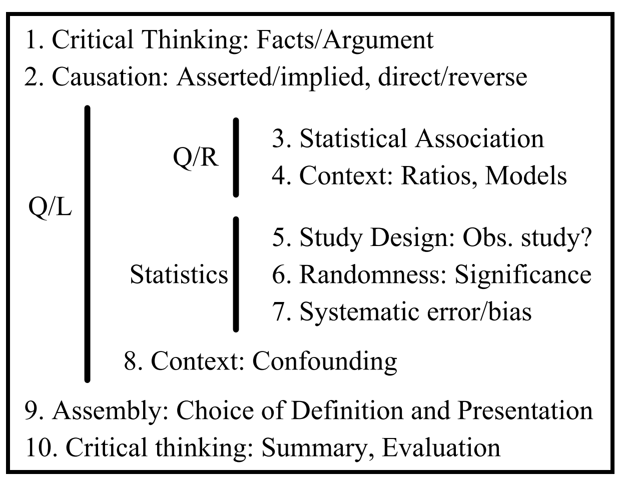

Figure 1: W. M. Keck Statistical Literacy Topics

Mathematical educators teaching a quantitative reasoning course might prefer a narrower, more focused approach that focused primarily on categories 3 and 4.

Statistical educators teaching a statistical reasoning course using textbooks by Utts (2004) or Moore (2008) might focus primarily on categories 5-7.

Educators teaching a data literacy or narrow quantitative literacy course might focus on 3-7 while those teaching an extended quantitative literacy or numeracy course might include 3 – 8. The latter is often the context for courses in epidemiology.

Educators teaching an argument-based numeracy or statistical literacy course might include all ten.

Teachers might be well-advised to start with their area of expertise and expand slowly outward. For example, mathematics and statistical educators might start with their area of expertise and extend their coverage to include both quantitative reasoning and statistical thinking (3-7). They could then move toward an extended quantitative literacy (3-8) by adding 8. A next step might be #9. Only when that base is very, very secure should they take on the critical thinking components of statistical literacy (1-10) by adding 1, 2 and 10.

7. Conclusion

This paper argues that students learning to be numerate need patterns they can practice on and master just as children learning to ride a bicycle need training wheels. Students need a structured approach to help them master the complexity of unstructured reality. This paper presents one approach that has been used by hundreds of students for over a decade.

But all of this is irrelevant. Those advocating a non-formulaic approach are completely justified in saying that a proper assessment of news-based numeracy should involve the student analyzing a news story without any prompting. This means that the final for a course should involve an unstructured assignment: an analysis without any specific prompting. Teachers using different approaches can see how successful their students were in this kind of final. Based on their students success (or failure), they can decide which method to use.

8. Future Work

More analysis is required to see:

- if the questions or categories are exclusive, exhaustive and essential.
- if there are elements in the Gal-Bracey-Lutsky approaches that are not adequately represented in the Schield approach.
- how useful is the Schield approach in dealing with a wide variety of news stories.

More data is needed to support the claim that a general formulaic approach:

- can avoid Madison's concern that such courses degenerate into a techniques-based algebra-like course and that students in such courses may not become fully numerate.
- can avoid Steen's concern that concept-based courses tend to be difficult for others to teach and for students to learn.

9. Acknowledgments

This work was performed as part of the W. M. Keck Statistical Literacy Project "to support the development of statistical literacy as an interdisciplinary curriculum in the liberal arts." This talk was presented at the 2008 meeting of the National Numeracy Network at Colby-Sawyer College. Thanks to Lynn Steen and Neil Lutsky for their suggestions.

10. References

Best, Joel (2008a). Birds – Dead and Deadly: Why Numeracy Needs to Address Social Construction. *Numeracy: Advancing Education in Quantitative Literacy*.[3]

Best, Joel (2008b). Pedagogical Challenges of Quantitative Literacy. *Calculation and Context.* Mathematics Association of America. 2323-2328.

Bracey, Gerald (2006). *Reading Educational Research: How to Avoid Getting Statistically Snookered.* Heinemann

Burnham, Thomas V. V. (2003). *Purpose and Nature of Statistical Literacy.* W. M. Keck Statistical Literacy Project Study Group at Augsburg College.

Gilman, Richard, Ed. (2006). *Current Practices in Quantitative Literacy.* Mathematics Association of America.[4]

Gal, Iddo (2000): Statistical literacy: Conceptual and instructional issues in D. Coben, J. O'Donoghue, & G. FitzSimons, (Eds.), *Perspectives on Adults Learning Mathematics* (pp. 135-150). London: Kluwer Academic Publishers.

Haack, Dennis (1976) "A Non-symbolic Statistics Course" in *Communications in Statistics*, Methods and Techniques.[5]

Haack, Dennis (1979) *Statistical Literacy.* Duxbury Press.[6]

[3] http://services.bepress.com/numeracy/vol1/iss1/art6/
[4] www.maa.org/EbusPPRO/Bookstore/tabid/37/Default.aspx
[5] www.StatLit.org/pdf/1976HaackCommInStatistics.pdf
[6] www.StatLit.org/Haack.htm

Haack, Dennis (1980) A Note on Teaching Statistical Literacy. *Teaching Statistics*, Vol. 2.1, pp 22-23. [7]

Lutsky, Neil (2006). Quirks of Rhetoric: A Quantitative Analysis of Quantitative Reasoning in Student Writing *Proceedings of the ASA Section on Statistical Education*, pp. 2319-2322.[8]

Madison, Bernard L. (2006). Pedagogical Challenges of Quantitative Literacy. *Proceedings of the ASA Section on Statistical Education*, pp. 2323-2328.[9]

Moore, David and William Notz (2008): *Statistics: Concepts and Controversies*. W. H. Freeman; 7th Ed.

Schield, Milo (1998). *Statistical Literacy: Thinking Critically about Statistics.* Inaugural issue of the Journal "Of Significance" produced by the Association of Public Data Users (APDU). [10]

Schield, Milo (2007). *Statistical Literacy: Teaching the Social Construction of Statistics.* Midwest Sociological Society.[11]

Schield, Milo and Cynthia Schield (2007). *Numbers in the News: A Survey.* Proceedings of the ASA Section on Statistical Education, pp. 2323-2328.[12]

Snell, J. Laurie (1999). *Using Chance Media to Promote Statistical Literacy.* Presented at the 1999 meeting of the American Statistical Association, Section on Statistical Education.[13]

Steen, Lynn (2004). *Achieving Quantitative Literacy: An Urgent Challenge for Higher Education* Mathematics Association of America.

Utts, Jessica (2004). *Seeing Through Statistics*. Duxbury Press; 3rd Ed.

Appendix A: Haack's Approach

Haack (1979) wrote Statistical Literacy: a textbook designed "to teach students how to interpret statistics so they can detect statistical doublespeak in the media and in their fields of study."

Haack (1980) noted that "Grading students is always a problem—more so it seems with an interpretive approach to teaching statistics. Three sources of information are being used in arriving at a student's grade in the course I am presently teaching.

- First, students are given weekly quizzes... Questions come from such local sources as the newspapers." "What might have influenced these statistics besides the selection of a representative sample...?" "I don't want students to simply list potential problem areas for any survey (as the source, the target population, wording of questions, timing, non-response rate, sample size and design, and method of contact) but to specify the areas of greatest concern for the specific example under consideration." "Here I would prefer that students question how a "child" is defined; in this example a child is anyone from age 2 to 20!"
- "Besides taking weekly quizzes, students also write (1/2-page) critiques of newspaper articles (5 per 15-week semester is sufficient). Again, one needs to look for some indication that the student can apply the principles discussed in class to the newspaper (or newsmagazine) example."
- "The third and final component of a grade requires that a student "dig deeply" into an application of statistics that the student has read or heard about. Students are encouraged to look for an example in their major field. For example, a project for a nursing student might involve critiquing an article in a medical journal. The purpose of this exercise is to get students to both look into a statistical application in their field of study and to go to an original source rather than feel they must always rely on the media's interpretation of statistics."

[7] www.StatLit.org/pdf/1980HaackTeachingStatistics.pdf
[8] www.StatLit.org/pdf/2006LutskyASA.pdf.
[9] www.StatLit.org/pdf/2006MadisonASA.pdf.
[10] www.StatLit.org/pdf/1998SchieldAPDU1.pdf
[11] www.StatLit.org/pdf/2007SchieldMSS.pdf
[12] www.StatLit.org/pdf/2007SchieldASA.pdf.
[13] www.StatLit.org/pdf/1999SnellASA.pdf.

Appendix B: Madison's Thoughts on QR/QL

"Based on [the] mathematical and statistical reasoning required to analyze and criticize various newspaper and magazine articles", Madison (2006) identified 17 conjectures involving QR/QL. These have been grouped into four categories.

Nature of QR/QL

1. QL is a habit of mind rather than a content-rich academic discipline.
2. QL requires practice beyond school.
3. Assessment of QL requires authentic situations.

Problems with QR/QL

4. High quality, effective curricular materials are scarce and scattered.
5. Abstracting generalities from contextual examples is difficult pedagogy.
6. Learning goals for QL are elusive.
7. Developmental levels of QL are neither understood nor articulated.
8. Performance standards for assessment are not established.
9. Multiple contexts challenge QL faculty and student understanding and knowledge.

Problems with student beliefs and attitudes

10. Students believe that QL is mathematics and behave as they do in traditional mathematics courses.
11. Students expect template problems and homework exercises that match the template, and template problems are antithetical to QL.
12. Students believe QL is mathematics and therefore deem it not relevant to their lives and set apart from other areas of study.

QR/QL News-Based Course Design

13. Course material must be fresh and engaging.
14. Excursions into political and social issues are sometimes delicate and mysterious.
15. Mathematical and statistical concepts occur repeatedly and unpredictably.
16. Use of technology is essential but often foreign to students.
17. Mathematics and statistics encountered is usually elementary.

As a result of this unstructured, problem-driven approach, Madison noted that "The students must be engaged in the material to a significantly larger extent than they are engaged in traditional mathematics or statistics courses." Madison concluded, "Keeping course material fresh offers a new challenge of producing textbook materials. Currently, my thinking is that the best I can do is to provide a skeletal framework for the course in terms of a textbook and require that the framework be filled out with fresh news materials."

Appendix C: Gal's News-Story Analysis Questions

Gall (2000) identified 10 types of questions that are relevant to analyzing statistics and chance in the news:

1. Where did the data (on which this statement is based) come from? What kind of study was it? Is this kind of study reasonable in this context?
2. Was a sample used? How was it sampled? Is the sample large enough? Did the sample include people/things which are representative of the population? Overall, could this sample reasonably lead to valid inferences about the target population?
3. How reliable or accurate were the measures used to generate the reported data?
4. What is the shape of the underlying distribution of raw data (on which this summary statistic is based)? Does it matter how it is shaped?
5. Are the reported statistics appropriate for this kind of data, e.g., was an average used to summarize ordinal data; is a mode a reasonable summary? Could outliers cause a summary statistic to misrepresent the true picture?
6. Is a given graph drawn appropriately, or does it distort trends in the data?
7. How was this probabilistic statement calculated, and are there enough credible data to justify such an estimate of likelihood?
8. Overall, are the claims made here sensible? Are they supported by the data? (e.g., confusing correlation with causation)
9. Should additional information or procedures be made available to enable me to evaluate the sensibility of these arguments? Is something missing?
10. Are there alternative interpretations for the meaning of the findings, different explanations for what caused them, or additional or different implications?

Appendix D: Bracey's Principles of Data Analysis

Bracey (2006) identified 32 principles[14] of data analysis that should be used when analyzing social science data. Schield grouped these into eight categories:

Facts

1 Do the arithmetic
2 Show me the data
5 Be sure the rhetoric and the numbers match.

Assembly/Social Construction

3 Look for and beware of selectivity in groups
8 Making certain you know what statistic is being used when someone is talking about the "average."
9 Be aware of whether you are dealing with rates or numbers. Similarly, be aware of whether you are dealing with rates or scores.
11 Be aware of whether you are dealing with ranks or scores.
16 Learn to be "see through" graphs to determine what information they actually contain.
22 Any attempt to set a passing score or a cut score on a test will be arbitrary. Ensure that it is arbitrary in the sense of arbitration, not in the sense of being capricious.
28 Make certain that descriptions of data do not include improper statements about the type of scale being used. For example "The gain in math is twice as large as the gain in reading."

Context

4 When comparing groups, make sure the groups are comparable
10 When comparing rates or scores over time, make sure the groups remain comparable as the years go by.
12 Watch for Simpson's paradox.

Explain

6 Beware of convenient claims that, what ever the calamity, public schools are to blame.
7 Beware of simple explanations for complex phenomena.
23 If a situation really is as alleged, ask, "So what?"

Randomness

13 Do not confuse statistical significance and practical significance.

Cause

14 Make no causal inferences from correlation coefficients.
15 Any two variables can be correlated. The resultant correlation coefficient might or might not be meaningful.
25 Rising test scores do not necessarily mean rising achievement.

Bias

19 A norm-referenced standardized achievement test must test only material that all children have had an opportunity to learn.
21 Scores from standardized test are meaningful only to the extent that we know that all children have had a chance to learn the material which the test tests.
31 In analyzing test results, make certain that no students were improperly excluded from the testing.

Tests

17 Make certain that any test aligned with a standard comprehensively tests the material called for by the standard.
18 On a norm-referenced test, nationally, 50 percent of students are below, by definition.
20 Standardized norm-referenced tests will ignore and obscure anything that is unique about a school.
24 Achievement and ability tests differ mostly in what we know about how students learned the tested skills.
26 The law of WYTIWYG applies: What you test is what you get.
27 Any tests offered by a publisher should present adequate evidence of both reliability and validity.
29 Do not use a test for a purpose other than the one it was designed for without taking care to ensure it is appropriate for the other purpose.
30 Do not make important decisions about individuals or groups on the basis of a single test.
32 In evaluating a testing program, look for negative or positive outcomes that were not part of the program. For example, are subjects not tested being neglected? Are scores on other tests showing gains or losses?

[14] http://www.statlit.org/Bracey.htm

Appendix E: Lutsky's Foundational QR Questions

Lutsky (2007) identified 10 foundational QR questions[15] that are relevant in analyzing numbers and data.

1. What do the numbers show? (a) What do the numbers mean? (b) Where are the numbers? Is there numerical evidence to support a claim? What were the exact figures? How can seeking and analyzing numbers illuminate important phenomena? (c) How plausible is a possibility in light of back of the envelope calculations?

2. How representative is that? (a) What's the central tendency? "For instance is no proof." Mean, Mode, and Median. (b) Interrogating averages: Are there extreme scores? Are there meaningful subgroups? Who's in the denominator? What's the variability (standard deviation)? (c) What are the odds of that? What's the base rate?

3. Compared to what? (a) What's the implicit or explicit frame of reference? (b) What's the unit of measurement? (c) Per what? (d) What's the order of magnitude? (e) Interrogating a graph: What's the Y-axis? Is it zero-based? Does it K.I.S.S., or is it filled with chart junk?

4. Is the outcome statistically significant? (a) Is the outcome unlikely to have come about by chance? "Chance is lumpy." Criterion of sufficient rarity due to chance: $p < .05$. (b) What does statistical significance mean, and what doesn't it mean?

5. What's the effect size? (a) How can we take the measure of how substantial an outcome is? (b) How large is the mean difference? How large is the association? (c) Standardized mean difference: $d = (\mu 1-\mu 2)/\sigma$

6. Are the results those of a single study or of a literature? (a) What's the source of the numbers: PFA, peer-reviewed, or what? (b) Who is sponsoring the research? (c) How can we take the measure of what a literature shows? (d) The importance of meta-analysis in the contemporary world of QR.

7. What's the research design (correlational or experimental)? (a) Design matters: Experimental vs. correlational design. (b) How well does the design support a causal claim? (c) Experimental Design: Randomized Controlled Trials (RCT): Research trials in which participants are randomly assigned to the conditions of the study. Double blind trials: RCTs in which neither the researcher nor the patient know the treatment condition. (d) Correlational Design: Measuring existing variation and evaluating co-occurrences, possibly controlling for other variables. Interrogating associations (correlations): Are there extreme pairs of scores (outliers)? Are there meaningful subgroups? Is the range of scores in a variable restricted? Is the relationship non-linear?

8. How was the variable operationalized? (a) What meaning and degree of precision does the measurement procedure justify? (b) What elements and procedures result in the assignment of a score to a variable? What exactly was asked? What's the scale of measurement? (c) How might we know if the measurement procedure is a good one? Reliability = Repeated applications of the procedure result in consistent scores. Validity = Evidence supports the use to which the measure is being put. (d) Is the measure being manipulated or "gamed"? The iatrogenic effects of measurement.

9. Who's in the measurement sample? (a) What domain is being evaluated? Who's in? Who's not? (b) Is the sample from that domain representative, meaningful, and/or sufficient? (c) Is the sample random? (d) Are two or more samples that are being compared equivalent?

10. Controlling for what? (a) What other variables might be influencing the findings? (b) Were these assessed or otherwise controlled for in the research design? (c) What don't we know, and how can we acknowledge uncertainties?

[15] http://serc.carleton.edu/quirk/CarletonResources/10questions.html

Appendix F: Burnham's Statements on the Purpose and Nature of Statistical Literacy

In 2003, the W. M. Keck Statistical Literacy Project investigated the purpose and nature of statistical literacy. The following was submitted by Tom Burnham.

Purpose of statistical literacy: To improve the quality of the student's decisions about issues for which statistical information is available.

Two main categories of issues: personal (e.g., health) and citizenship (e.g., voting).

Definition of statistical literacy: The habit of noticing without specific prompting **the strengths and weaknesses** of such claims and reports of statistical information and arguments based thereon as commonly appear in the non-technical media.

- "the habit": NOT just the ability – use it regularly in real life or lose it. Students need to do more than pass an exam; they need to build durable habits. See Moore on the non-innateness of the modes of thought of the liberal art and of stats; and Chance on developing habits and on the need to reinforce them in follow-up courses. Building the habit of statistical literacy will involve replacing any old habits which get in the way. See Garfield's examples of errors which students need to be taught to avoid.
- "... of noticing without specific prompting": if the student requires prompting about particular issues or aspects, then he doesn't have a durable habit and the impact on his life will be minimal. The ideal final exam for a statistical literacy course is about a half dozen short articles from the media, each followed by the single word "Comment." Any more detailed instructions or questions will trigger recognition rather than recall and will not test firm habitual knowledge and skills. It is regrettable that grading such a test is expensive, but I doubt that anything less will prove that the above-stated purpose has been fulfilled.
- "the strengths and weaknesses": that noticing strengths and weaknesses will improve the quality of decisions seems self-evident; the question is: which strengths and weaknesses should we expect students to learn to notice? Part of the answer is provided by the source material: statistical concepts and techniques which are hardly ever mentioned in the non-technical media may reasonably be excluded. Excluded items might include such things as stratified samples and specialized significance tests but not case-control studies and cohort studies. Border line items might include p-value. Basically, "strengths" refers to those aspects of sound statistical reasoning which are commonly found in media reports or which are normally associated with the terminology of such reports; and weaknesses includes the absence of strengths. Many more weaknesses will be omissions (not necessarily malicious) from the media reports of details and/or qualifications which appeared in the original report. Students need to know that certain types of arguments gain strength from the combinations of pieces of information, and they need to notice whether all of the pieces are reported. E.g., centers often need spreads, and associations usually need margins of error or confidence intervals.

I believe that the previous three items together presuppose an adequate degree of understanding, interpretation and reasoning, with one major exception: students need to be taught that certain words which sometimes appear in media reports are used in technical senses which differ substantially from the common senses of those words; e.g., explain, account for, control.

- "claims": does anyone doubt that fabricated, erroneous and/or excessively abbreviated reports are common enough to need to be recognized and dealt with?
- "reports of statistical information": the obvious subject
- "and arguments based thereon": the frequency of arguments which urge some action (personal or social) on the basis of cited statistics hardly needs comment. The student needs to recognize the difference between statistical conclusions and action decisions, and think clearly about whether the former is sufficient warrant for the latter. See Garfield on "the outcome orientation"; the conversion to a binary decision is not statistically valid but is involved in many real world decisions.
- "commonly appear in the non-technical media": a reasonable upper limit which (we hope) is attainable by most college students at an affordable cost.

Valuable but not obviously affordable additional issues to which (IMHO) students should be sensitized:

- Measurement issues, e.g., objective vs. subjective; direct vs. proxy.
- Common confounders, e.g., Hawthorne effect, placebo effect.
- Sampling problems. e.g., convenience sample, nonresponse bias.

A valuable but expensive exercise: have students evaluate a media report, then look up the original source and compare that with the media report.