CONNECTING WITH REALITY

A learning-theory analysis of the successful reform of the elementary statistics curriculum

Hunter Ellinger (hunter.ellinger.org, ellinger@io.com)

Introduction:

Successful curriculum reform, while clearly needed in many instructional areas, is sufficiently rare to deserve careful attention when it does happen. This paper traces the origin and nature of the changes in the curriculum used in introductory statistics courses. In addition to discussion of the elements that have contributed to the success of this particular change, an attempt is made to derive some general conclusions that can be used in other areas where a closer connection between teaching and real-world activities is desirable.

The pattern of this change is analyzed in terms of several theories of learning that have been used over the last few decades. The primary analysis is of the implications of learning theory for the content and methods of the introductory statistics course. But a secondary analysis is made of the "learning" involved within the community of educators during the effort. After all, reform is (or should be) a learning experience for all involved.

Historical Background:

The relationship of statistics to academic teaching is complex. Statistics has been something of a foster child, with primary custody passing from the natural sciences to the social sciences and ending up, somewhat grudgingly, under mathematics. This in part reflects the fact that statistics is a "new" discipline, which has only received serious academic attention in the late 1800's and still has sectors in which the empirical practitioners use techniques that are more advanced and complex than those recognized and mastered by the theory-based academic community. But the lack of clarity of statistics' academic home also reflects its threefold origin: physical science (assessing and refining measurement accuracy), social science (concise data description), and mathematics (the implications of probability). The emphasis of statistics education has wandered between these three poles, only recently making a serious attempt at unifying the strands.

Although the foundations of mathematical statistics had been completed soon after the turn of the century, statistics education at that time focused almost exclusively on what would now be called descriptive statistics. Issues of measurement were dormant, with those of the physical sciences reduced to cookbook formulae and those of sampling theory seen as too advanced and abstract for the undergraduate curriculum (and usually neglected in practice as well). This is a period during which statistics was seen primarily as an aspect of economics or sociology, and the emphasis in courses was on methods of assessing, translating, and summarizing data.

For example, a 1890 article in the *Journal of the American Statistical Association* (Vol. 2, pp. 83-86) on "The Study of Statistics in Colleges" by F. A. Walker (then ASA President), listed the "three uses of statistical study" as "to detect the fallacies in conclusions drawn by others from quantitative statements", "to enable the writer or speaker upon politics, economics, history, or sociology safely and effectively to illustrate and emphasize his conclusions", and "for the discovery of social laws". Similarly, a description of "A University Course in Statistics" in the next volume (p. 186, referring to Fisher's first class at the University of Chicago) lists the main activities planned as "establishment of statistical offices and their organization, collection and elaboration of data, detection and elimination of errors, presentation of results in tabular form, and training in graphic representation".

By 1926, ASA President Robert Chaddock, in his Presidential address "The Function of Statistics in Undergraduate Training" (*JASA*, Vol. 21, pp. 1-8) gives a quite different emphasis. Noting the need to move statistical training from a graduate-school add-on to the role of a foundation course for lower-division students, he lists the following set of "most important services which elementary statistical training can render":

- developing the "critical judgement of the student",
- emphasizing the "essentials of the scientific method",
- help the student "realize the interdependence of different factors",
- "supplementing deductive processes",
- observing the "stability in results obtained from the analysis of large numbers of cases", and
- "preserving useful mathematical concepts by showing their application to social data".

These words imply that statistics is about *thinking* rather than just counting or figuring -- they could fit into a current "reform" statistics textbook without change.

Mathematics captures the curriculum:

But as the efforts to offer statistics training to lower-division students bore fruit, the loss of the graduate-level context created pressures contrary to these intellectual goals. Students who lacked deep subject understanding, and who were gathered from different fields (or had not yet chosen a field), were not well-positioned for thoughtful work with real data. This problem was exacerbated by progress in sampling theory, which now extended to social-science statistics the high-powered mathematical methods originally derived by Gauss and others in the nineteenth century for astronomical measurements. The combination of more math with less real data led to a situation which the Committee on Applied Mathematical Statistics of the National Research Council described in 1947 as:

"Often the main objective of such laboratories [in elementary statistics classes] is the calculation of means, variances, correlation coefficients, and other statistical quantities from numerical data of various types. The laboratory work should place more emphasis on interpreting or drawing inferences from data and on the nature of those inferences. Simple experiments should be devised for illustrating probability laws of various kinds, for carrying out sampling operations and other random processes. The traditional flipping of coins and rolling of dice are not adequate for illustrating many important random processes. The mathematical theory of many of these processes is too complicated to be handled at the elementary level; their experimental demonstration will give the beginning student some feeling of their significance."

Note that even this critical view does not see using real-world data as a feasible approach -- it just calls for experiments that better mimic the complexity of the world. The objection is not to abstraction *per se*, but to coin-flip oversimplicity. Clearly expectations had shifted. For example, even in the 1948 article which cites the above quote approvingly (George Snedecor," A Proposed Basic Course in Statistics", *JASA*, Vol. 43, pp. 53-60) the 12 topics proposed for a one-semester introductory course include only one (Non-normal Distributions) that is not focused on sampling or regression theory. [The outline for that topic does give the advice: "*Emphasize that no known distribution is normal*", but does not follow up the implications of that fact.]

That something important was being omitted from introductory statistics courses was not totally overlooked. In his 1950 Presidential address to the ASA ("Undergraduate Statistical Education", *JASA*, Vol. 46, pp. 1-18), S. S. Wilks says that he thinks that with "careful selection of material" the essentials for introduction to statistics could be boiled down into a sequence of two full-year courses. Thus his era did not deny the value of authentic data analysis, just as the current reformers do not deny the value of knowledge of probability and hypothesis testing (reform guru David Moore says "You can't be too rich or too thin or know too much mathematics"). But to make a single-semester course out of topics that would crowd two years, one must declare priorities -- this calendar-based fact is the basic source of the curriculum battles.

<u>Transition to recent times:</u>

While the mathematical component came to dominate the introductory statistics curriculum, there were some disturbing problems. The gap between academic instruction and practicing statisticians was widening, with many important (and, in retrospect, essentially correct) applied techniques being deployed with little if any theoretical support. The very success of subsectors such as statistical quality control separated them from the educational mainstream into a training context, at least as far as methods were concerned. Some practicing statisticians (most notably W. E. Deming) criticized the formal inference methods taught in schools as being inapplicable to most of the cases in which they were used.

The mathematical tools that were feasible to use (necessarily by hand at this period) were based on assumptions of limited applicability, but still formed the bulk of the curriculum. Even so, the mathematics involved seriously challenged the ability of typical lower-division students, most of whom did well to learn the mechanics of statistical calculation and problem recognition, with little chance of building a serious mathematical understanding of the formulae and no chance of understanding the statistical rationale as well.

The result of this was that for introductory students the loss of real-world connection to the problems of a specific discipline was not compensated for by insight into the source and generality of the mathematics involved. If it was true that "statistics has been captured and enslaved by mathematicians" (Oscar Kempthorne, "The Teaching of Statistics: Content versus Form", *The American Statistician*, Vol. 34, No. 1, p. 21), the captive was displaying the typical slave's combination of lethargy and resentment -- with the potential for rebellion if the balance of power changed.

The best that could be achieved prior to that revolution was to organize the conditions of servitude competently and smoothly, which in this context meant a curriculum and set of instructional designs which brought students to basic competence in the recitation of definitions and in the application of statistical techniques to routine problems drawn from a well-defined set of patterns. This became the goal of authors of introductory textbooks, with Mario Triola's *Elementary Statistics* (Benjamin/Cummings Publishing, Menlo Park, California) gaining a leading position by the 1980s by skillful execution of this agenda.

Relaxation of rigor in derivation of formulae permitted Triola to mention a great variety of concepts and techniques, but the effect typically produced was broad superficial acquaintance rather than authentic mastery, with its awareness of the limitations of the recipes. Almost all of the work posed for the students in this era can be better classed as exercises than as problems; they were never called upon to struggle with the issues raised by real-world data. The resulting experiences could contribute little to either statistical or mathematical maturity, undercutting support for the curriculum from the academic research-statistics communities.

Since the practitioner community was already alienated and the mainstream mathematicians had never really respected the field, this left the elementary-statistics curriculum with very few friends in high places. This greatly reduced the obstacles to the reform movement when it emerged.

One revolution leads to another:

Conforming neatly to dialectical theory, the most recent revolution in statistics teaching resulted from the great changes in the means of production (i.e., computation) and distribution (i.e., communication) following from progress in computer technology. As David Moore, who has taken over Triola's position as the leading textbook author (and is a recent President of the ASA), explained it: "Statistics has moved away from mathematics back toward its root in scientific inference and the analysis of data. ...The most important driving force in this shift of emphasis is the computer revolution." (Quoted by George Cobb in "Teaching Statistics" in *Heeding the Call for Change: Suggestions for Curricular Action*, The Mathematical Association of America, 1992, p. 4.)

The new curriculum has a new set of emphases. In Moore & McCabe's bestselling textbook *Introduction to the Practice of Statistics* (Freeman, New York, 1989), for example, probability is not introduced until chapter 5 (of 11). It is preceded by three chapters on "Looking at Data" (distributions, change & growth, and relationships), and one on "Producing Data". While this approach is reminiscent of the turn-of-the-century orientation toward critical enumeration, the new ground is indicated by that fourth chapter on production of data, which provides a basis for discussion of the scientific method of "controlled" experiments and of the need to take an active role in *producing* informative data rather than just passively making the best of muddled numbers produced by someone else for an incompatible purpose. One of Moore's quotes (p. 6, "What is Statistics?" from *Perspectives in Contemporary Statistics*, MAA Notes Number 21, Math. Assn. of America, 1992) communicates his point: "Good data are as much a product of human skill as hybrid corn and compact disc players."

When the material on probability theory is finally introduced, it is not allowed to take over. Moore again: "In this setting, we should limit ourselves to those aspects of probability that are essential for inference. We cannot afford to give probability its due as a subject in its own right." The fact that an explicit or implicit mathematical model of the process which produced the data must underlie any use of that data for inference is stressed, with some applications that are not amenable to straightforward inference computation (e.g., continuous-process control charts, whose measurements often lack full independence) introduced early enough to show the need for careful thought and some restraint in conclusions.

Even statistical theory is used sparingly. Moore gives explicit arguments against the unnecessary inclusion of theories that depend on assumptions (such as normality of the data or equality of the variance of different sampling procedures) that may be difficult to ascertain or verify, or theories (e.g., Bayesian inference) whose demands on students' subtlety of distinction are excessive. The focus is on statistical concepts (variation, data, data production, quantification, explanation, correlation, etc.), with only the high points of the theoretical conclusions that can be constructed using these concepts.

Moore concludes his exposition of the motivation for his approach by saying "**The higher goal of teaching statistics is to build the ability of students to deal intelligently with variation and data**." (MAA Notes #21, cited above). The goal here is thus development of higher-level intellectual skills (a set of "cognitive strategies"), not possibly-unthinking mastery of a fixed set of procedures. It is probably the success the statistics-reform group had in articulating this way of looking at things which has protected it from the "dumbing down" accusations levied against the subsequent calculus-reform movement.

What can be learned from learning theory about learning about theory:

The use of a learning-theory concept ("cognitive strategy") in the above analysis is significant on more than one level. First, the nature and changes in the statistics curriculum are illustrations of the application of explicit or implicit learning theories. Also, the controversy had echoes of the general battles about learning theory (the parallels are all the stronger because psychology and statistics became academic subjects at about the same time). Finally, the outcome can be fruitfully analyzed in terms of learning theories -- why was the reform message "learned" in this case when similar messages (articulated by the same people in many cases) have been much less successful in other areas?

That the reform was presented in part as an aspect of the constructivist analysis was natural, given its timing. But in addition to any fashionableness, this instructional area is a particularly good fit to many constructivist ideas. Most students have a set of natural, but logically incoherent, ideas about probability and inference from data. These are sufficiently deeply rooted that they must be taken into account in teaching, with appropriately-chosen direct experiences being much more effective than theory in changing people's minds. Further, since the reform-statistics goal is producing competence in deploying cognitive strategies for turning data into information, *enabling active construction of meaning by students is the central point of the course, not just an instructional tactic.*

The leading member of the statistics-reform group from an educational-psychology background has been Joan Garfield. Her 1988 article (with Andrew Ahlgren) on "Difficulties in Learning Basic Concepts in Probability and Statistics: Implications for Research" (*Journal for Research in Mathematics Education*, Vol. 19, Number 1, pp. 44-63) both summarizes the history of research and gives a set of recommendations:

" Teachers should:

1. introduce topics through activities and simulations, not abstractions;
2. try to arouse in students the feeling that mathematics relates usefully to reality and is not just symbols, rules, and conventions;
3. use visual illustration and emphasize exploratory data analysis;
4. teach descriptive statistics alone without relating it to probability;
5. point out to students common misuses of statistics (say, in news stories and advertisements);
6. use strategies to improve students' rational number concepts before approaching proportional reasoning;
7. recognize and confront common errors in students' probabilistic thinking;
8. create situations requiring probabilistic reasoning that correspond to the students' views of the world. "

While generally supportive of these recommendations, Garfield points out that "little empirical research has focused on the effectiveness of the different instructional methods or teaching approaches in developing statistical and probabilistic reasoning". This lack of data would seem to enhance the importance of using learning theory to guide curricular development in this area. In fact, in the reform-statistics spirit of taking responsibility for producing useful data, learning theory should be used to help decide what empirical research is needed.

The relevance of several theoretical approaches can be seen. There has been some childhood-oriented work by Piaget and Inhelder (*The origin of the idea of chance in children*, London, Routledge & Kegan Paul), but the fact that the difficulties are not developmental in the normal sense is indicated by reports (from the second NEAP mathematics assessment, reported on page 50 of Garfield's JRME article) that "students' intuitive notions of probability seemed to get stronger with age but were not necessarily more correct". There may be genetic biases; Garfield goes on to say (page 58): "The misconceptions already identified are not isolated errors of information or arbitrary habits of thinking. It seems, rather, that misconceptions are a way of thinking about events that is deeply rooted in most people, either as learned parts of our culture or (in the extreme) even as brain functions arising from natural selection in a simpler time."

Note that in this field even avowed constructivists such as Garfield show none of the "stay out of the way and let the kids figure it out for themselves" complacency which is seen in some other areas. Instead, she calls on teachers to "confront naive intuition directly" -- deliberate instruction is clearly going to have to play a leading role in statistics education.

The dependence of statistics on real-world activities for both motivation and the data that is the basis of authentic and effective instruction makes it a natural place for an analysis based on the society-oriented ideas of the classic Russian educational psychologist Lev Vygotsky. This is clearly a discipline in which the world is driving both theory and learning, not the other way around, and learning clearly benefits greatly from social interactions both between students and the outside world (by active production or collection of data) and among students (which promotes constructive articulation of ideas and conclusions). Vygotsky's view of theory (especially appropriate *concepts*) as a generative attractor in learning (but as a complement, not substitute, for experience) is probably even better matched to the needs of statistics than current fashions acknowledge.

The more recent theoretical developments in theories of learning and instruction also provide some insights into the statistics-curriculum controversies. The recognition of different levels of intellectual abilities is very helpful, especially the distinction between skills, procedures, and cognitive strategies. The principle that the establishment of dissonance is needed prior to instruction provides a basis for criticism of the hodgepodge of topics and exercises common a generation ago, since they were given more as miscellaneous facts than as solutions to something the student was puzzled about. In the reform approach, dissonance is produced by having the students interact with data and simulations.

The work on the impact of affective issues on learning also has clear relevance to statistics education, since many students have an emotional aversion to statistics even greater than that for mathematics in general. Much of this is rooted in issues best considered in the light of attribution and efficacy theory. The statistics-reform group's success with methods that avoid abstraction and use real-world-related participatory activities is explained in part by the changes such an approach makes in students' beliefs about their ability to participate successfully in the dreaded math classes.

While little use has been made of it yet, the role modeling analyzed by social-cognitive theory could also be used to good effect in statistics instruction. (Currently, it probably works the other way -- green eyeshades aren't cool.) What could be added would be showing how statistical thinking is used by in-charge decision-makers who are the masters (and producers) of data rather than the servants of it. The image of CEO rather than clerk should be projected, along with some collateral stories of scientists who have made important discoveries by careful attention to data (rather than by genius or luck). The case-study approach by which data is usually provided gives a natural context for introducing such models.

Learning about efforts to reform learning:

While the role-model approach is not yet used much in instruction of students, it is probably the largest single factor contributing to the success of the educational establishment "learning" the lessons of this reform movement. The heroes of statistics were at least allies, and often champions, of reform. Even those numerous teachers of statistics classes whose background is in mathematics rather than statistics are awed by the renown of a Deming, and can see that his experience and recommendations are worth attention. Statistics was fortunate in having all the celebrities on one side; a major source of the continuing fight in calculus reform is that there are at least two important groups of role models (engineer/applications and scientist/ research) with differing goals.

Beyond the impact of individual models, however, the statistics reform effort was strengthened by being well connected to many parts of the world. This gave it (as well as more interest to the students) allies both within and outside academic circles. The political power of being aligned with the needs of the world is often greatly underestimated, and ensuring that such an alignment exists and is communicated to those affected is essential for successful revolutions of any kind.

Any reform movement starts with its members first discovering dissonance for themselves, and then preaching it to the world -- "Things aren't working; you have nothing to lose but your chains." The strength of this strategy is that it attracts all who are discontent with things-as-they-are; the danger is that these people will not necessarily agree with each other about what new should come. And it is hard to replace something (even something bad) with nothing. This is why a period of exploratory course and textbook development is essential prior to the real battles. (A corollary is that there is always useful work to do for the next revolution, even when no major public battles are underway.)

It is thus really the institutional-efficacy issue that is the most effective defense of an old regime. "We can't do anything better at this level" [or with these students, or with our budget] is an effective response (especially when it is true, in which case the fight may need to be shifted elsewhere). In addition to attacking this issue by demonstrations of what in fact can be done, reformers should also pay close attention to any changes in institutional conditions that could offer opportunities to force a general reassessment of the possibilities. For statistics, such an opportunity arose when widespread computer use both added tools for exploratory data analysis and removed the need for much of the computation-method training that had previously been seen as essential.

A final thought. Of all the learning-theory-analogous strategies for reform (in curricula or elsewhere), the one that seems to me to offer least possibility for success is individual free discovery. Educational reform, like human life in general, is an essentially social activity, in which any serious progress must be built on the achievements of previous generations.