

Evaluating Teaching Methods: Is It Worth Doing Right?

Robert G. Nelson and Norbert L. W. Wilson

Department of Agricultural Economics & Rural Sociology

Auburn University

Contact Info: Robert G. Nelson

Department of Agricultural Economics and Rural Sociology

Auburn University, AL USA 36849-5406

phone: (334) 844-5621

fax: (334) 844-5639

email: nelsorg@auburn.edu

*Selected Paper prepared for presentation at the Southern Agricultural Economics Association
Annual Meeting, Dallas, TX, February 2-6, 2008*

Copyright 2008 by Robert G. Nelson and Norbert L. W. Wilson. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Evaluating Teaching Methods: Is It Worth Doing Right?

Robert G. Nelson and Norbert L. W. Wilson

Abstract:

Reviewers of manuscripts on classroom experiments often ask the authors to provide evidence of the effectiveness of the method, presumably to justify substituting experiments for lectures.

After reviewing the current state of evaluation methodology, we argue that such evidence may be neither sufficient nor even necessary for the purpose.

It is easy to lose sight of the fact that the variables that influence the utility or production functions for teaching and learning in the classroom are highly endogenous and complex, and the phenomenon of learning involves multiple outputs in multiple domains (Becker 1982; Becker et al.). In one taxonomy addressing development in the *cognitive* domain (Bloom and Krathwohl), the hierarchy of learning outcomes progresses through knowledge, comprehension, application, analysis, synthesis, and evaluation. In another taxonomy involving the *affective* domain (Krathwohl, Bloom, and Masia), the hierarchy ascends through receiving, responding, valuing, organizing, and characterizing. We would argue that the few available standardized and nationally-normed tests – that is, those that have been shown to be valid and reliable – at best evaluate only the lower levels of the cognitive and affective domains, and would not be appropriate for upper-level curricula, such as capstone agribusiness courses.

Moreover, there are many other measures of performance and dimensions for assessment besides standardized tests. Direct measures include exams, papers, presentations, portfolios, competitions, and judged performances. Indirect measures include evaluations by students, instructors, peers, alumni and employers, using such instruments as surveys, interviews and anecdotes. Before an evaluation could be done in an agribusiness curriculum, a measurement instrument would need to be chosen. However, to our knowledge there is no consensus on teaching methods, learning outcomes, or standards of performance for agribusiness curricula, nor has there even been any consistent call for such standards in the literature.

Considerations in Evaluating Effectiveness of a Teaching Method

Anecdotal reports and informal surveys of teaching effectiveness, learning outcomes and student satisfaction are notoriously subject to biases and should be heavily discounted. Even

formal student evaluations and course grades present problems in inference, such as reliability, validity and sample selection (Becker 1997). A definitive study of the effectiveness of experiments in teaching agribusiness would need to address the same considerations identified by Fraas over a quarter-century ago: random sampling and random assignment of treatments; adequacy of research design; duration of student exposure to the technique; reliability and validity of the testing instrument; and interaction effects between student characteristics and methods of instruction. These are addressed in sequence below, in the context of agribusiness instruction.

Random sampling and random assignment of treatments: Because of size and scheduling constraints in agribusiness courses it is rarely possible to randomly assign students or instructors to different instructional treatments. Moreover, few agribusiness courses are large enough to split into a treatment and a control, even assuming there was a standardized, orthodox approach to teaching agribusiness that would be generally accepted as the appropriate control. The lecture-discussion format has standing only because it is the status quo.

Adequacy of research design: Aspects of design have been meticulously addressed in a few studies of the effectiveness of using experiments to teach microeconomics. Cardell et al. described an experiment involving all the introductory micro- and macroeconomics courses at Washington State University ($n > 1,400$) with a treatment of four market experiments compared to a standard lecture-discussion method as a control. A pilot semester was devoted to constructing syllabi, standardizing instructional techniques, developing policies for attendance, and designing methods for collecting personal data, such as time diaries to measure study effort. Students and instructors were randomly assigned to treatment or control, and these were paired by common days and times. Crossovers and dropouts were tracked, and adjustments for sample selection

bias were considered. Students were given both the micro- and macroeconomics parts of the multiple-choice Test of Understanding in College Economics (TUCE) at the beginning of the course, and again as part of the final exam. The dependent variable was the difference between the student's scores on the post-test and pre-test, which they termed "value-added." The authors offered a proof that this difference is unbiased and efficient, whereas a post-test score alone is not. An OLS model was used to regress value-added against a number of suites of variables and their interactions. The continuous variables included: GPA, SAT-math, SAT-verbal, ACT, age, attendance, class standing, and study hours. The discrete variables included: treatment, gender, previous economics course, currently in microeconomics course, ethnicity, source of GPA record, highest math course taken, missing scores, and extracurricular activities. Five models were reported, with adjusted R^2 s ranging from 0.09 to 0.11. None of the models showed any net positive impact due to the experimental treatment. Conversely, Dickie conducted a similarly rigorous study, but found a statistically significant increase in scores on the Microeconomics Section of the TUCE. He attributed the disparity in the results of the two studies to the emphasis on understanding microeconomic principles, rather than coverage of both micro- and macroeconomics as in the Cardell et al. design. But even these sophisticated designs may be missing the point. In summarizing efforts to use test scores to measure educational production functions, Becker (1997) commented:

...the use of the educational production functions with test scores as the only output measure is too narrow. Pre-and posttest, single-equation specifications, with potentially endogenous regressors, simply may not be able to capture the differences that we are trying to produce with diverse teaching methods.

Adjustments for sample selection problems are needed but even after these

adjustments with large samples, failure to reject the null hypothesis of no instructional effect may point more to deficiencies in the multiple-choice test outcome measure or application of the classical experimental design than to the failure of the alternative instructional method under scrutiny.” (p.1367)

Duration of student exposure to the technique: It is not readily apparent to us how one would address the question of how many experiments would be suitable -- or would constitute a valid treatment -- for a particular course. Certainly this would be influenced by how much economic theory the instructor preferred to incorporate. A single experiment would be unlikely to impact any learning performance criterion other than a tailored, special-purpose quiz. Gremmen and Potters, Beil and Delemeester, and Frank all used this approach. Other studies used multiple experiments as a treatment and a comprehensive exam as the performance criterion. For example, Cardell et al. used four experiments and the TUCE; Yandell used six experiments and the final exam; Dickie used seven experiments and the TUCE; Fraas used seven simulation-gaming activities (scheduled to account for 50% of class periods) and the TUCE; Mullin and Sohan used eight experiments and the TUCE. As a further complication for evaluating experiments that demonstrate economic theory in an agribusiness course, it is well-recognized that some experiments and protocols are better suited to some syllabus topics than others (Mullin and Sohan). To accommodate this concern, Dickie suggested administering a separate test after each experiment, as well as an aggregate exam.

Reliability and validity of the testing instrument: To our knowledge, no instrument for testing agribusiness knowledge has been evaluated for reliability or validity. Because agribusiness instruction does not have a standardized performance criterion analogous to the TUCE, differences between pre- and post-test knowledge cannot be consistently measured. In any case,

as noted earlier, multiple choice questions are not the only -- nor necessarily the best -- way to test the acquisition and retention of knowledge (Hansen), especially considering that performance is heavily influenced by learning styles. Herz and Merz addressed the dimension of learning styles by using a novel measure to evaluate the effectiveness of simulation games in economics. They asked students to rate their progress through Kolb's four phases of experiential learning ("behaviors"): concrete experience, reflection, conceptualization, and experimentation. Their 22-question instrument elicited responses concerning the degree to which the experimental treatment supported the student in applying and improving those specific behaviors, and included a range from "no support" to "very strong support." Results showed that the simulation game outperformed a conventional seminar in all four phases of the learning process, but was best at supporting the subcategories they described as "creative search," "active involvement," and "social interaction."

Interaction effects between student characteristics and methods of instruction: Fraas suggested that one could use an understanding of the interaction effects between student characteristics and methods of instruction to assign students to the course sections that would most improve their retention of economic knowledge or nurture their continued interest in the economics curriculum. He identified a number of such effects as significant, including pre-course knowledge, previous economic training, and SAT scores. Cardell et al. found attendance, math scores, GPA, ethnicity and courses in microeconomics to be significant, while Dickie added age, ACT, and exposure to previous experiments.

The extra effort required by the instructor to conduct experiments instead of lectures may be another, underrated, interaction. Furthermore, it is generally agreed that extra credit and cash incentives can significantly influence motivation, but what is less appreciated is that they may

also generate the interaction effect of increasing attendance rates. Becker (1997) noted that "...students do not take the (TUCE) test seriously when it does not count in the course grade. Yet, many educators continue to overlook the effect of incentives on measured student performance" (p. 1365).

Finally, to keep this discussion in perspective, all these factors would also be considerations in evaluating the relative effectiveness of any other teaching method, such as case studies or field trips, should a similar effort be made to examine them with this kind of rigor.

Cost/ Benefit Analysis

A better way to evaluate a new teaching method is cost/benefit analysis. A number of authors have focused attention on the costs and benefits of using experiments in the classroom. In their compilation of 113 non-computerized classroom experiments, Brauer and Delemeester developed a bi-matrix of costs and benefits to the student and the instructor that is useful in capturing many of the dimensions of effectiveness.

Among costs, most often mentioned is the opportunity cost of the additional time the instructor needs to develop, coordinate, administer and integrate an experiment into the curriculum. Another potential cost is the possible adverse effect on the instructor's confidence, authority or reputation when an experiment does not produce the expected result, or the classroom environment seems less organized or structured (King; Yandell).

Costs to the student may include reduced achievement among students with certain learning styles, personalities, aptitudes or motivations. In Fraas' study of undergraduates randomly assigned to either a lecture-discussion section or a simulation-gaming section of introductory economics, students with low pre-course knowledge, no previous economic

training, or low SAT scores performed better on the exam after exposure to the gaming treatment than the lecture. In contrast, Dickie found that while experiments improved learning of microeconomics principles among the majority of students, they did not benefit students whose GPA's were below 2.04 or whose ACT scores were below 16.6 (both about one standard deviation below the mean).

Among benefits, the vast majority of anecdotal evidence is that experiments are extraordinarily popular and successful in the classroom, and to our knowledge no authors have reported dissatisfaction with the results of their efforts to introduce the method in their classes. Empirical tests have consistently shown that students exposed to experiments on average do at least as well as those exposed to lectures, and often better (Brauer and Delemeester). Several other student benefits are frequently mentioned, such as more fun in class, more engagement, better attendance, greater attention and higher effort. However, inasmuch as these are inputs in the function that produces higher exam scores, including them separately on the benefit side of the ledger would seem to be double-counting. On the other hand, Becker (1997) notes that students who have saved time through more efficient learning in one subject may invest that time in other subjects, in part-time jobs or in leisure -- all contributing to a general increase in student welfare.

Brauer and Delemeester remark that a frequently-ignored quadrant in their bi-matrix is the benefit to the *instructor*. This can take many forms, including: pleasure in a more genial relationship with students (King); better student evaluations (Yandell); pedagogical efficiency by laying empirical foundations for understanding more elaborate concepts (Nelson and Beil 1994) or by substituting a single experiment for a number of lecture examples (Hauptert); sense of accomplishment in providing a connection between theories and key features of markets and

institutions (Smith); inspiration for generating and exploring research topics (Noussair and Walker; Becker and Kennedy); and reputation for projecting the image of a scientist (Barnett and Kriesel).

Even a simple bi-matrix such as this offers a far richer dimensionality to the question of introducing new teaching methods. Ultimately, each instructor will individually weight the perceived costs and benefits of the elements in each quadrant in order to decide whether to use a new method or not, and even then trial-and-error learning is likely.

Conclusions

To forestall any self-congratulation about progress in the evaluation of teaching methods in economics, we should note that in the early 1970's Wentworth and Lewis reviewed nearly 50 studies evaluating educational games and simulations and many of their criticisms still apply to studies done 30 years later. They questioned the credibility of studies that had not demonstrated the reliability or validity of their instruments. They lamented that teachers had not been selected randomly, and harbored "suspicions of possible developer bias." They censured studies that were "marred by a haphazard selection of participants," and stressed that "findings must be viewed cautiously because of the small sample used in the study." Generalizability was jeopardized for lack of control groups or independent replication, testing procedures and research designs were "inadequate," statistics were "unsophisticated" and the totality of research in the area of learning processes was "superficial and confusing." In that time and context, the authors might have opined that if an evaluation wasn't worth doing right, it wasn't worth doing at all. But given that today there is still no standard method of teaching agribusiness courses -- other than the status quo of the lecture-discussion method -- and given that there is no single, accepted,

valid and reliable testing instrument (nor perhaps should there be), we might well assert that if an evaluation isn't worth doing at all, it isn't worth doing right. Maybe Rendigs Fels described it best, and unintentionally resolved the circle of reasoning, when he titled his critique of experimental economics in the classroom "This is what I do, and I like it."

References

- Barnett, B.J., and W. Kriesel. "Agricultural Economists' Use of Classroom Economic Experiments." *J. Agr. and Applied Econ.* 35(August, 2003):321-335.
- Becker, W.E. "Teaching Economics to Undergraduates." *J. Econ. Lit.* 35(September 1997):1347-1373.
- ____. "The Educational Process and Student Achievement Given Uncertainty in Measurement." *Amer. Econ. Rev.* 72(March 1982):229-236.
- Becker, W., R. Highsmith, P. Kennedy, and W. Walstad. "An Agenda for Research on Economic Education in Colleges and Universities." *Amer. Econ. Rev.* 81(May 1991):26-31.
- Becker, W.E., and P.E. Kennedy. "Does Teaching Enhance Research in Economics?" *Amer. Econ. Rev.* 95(May, 2005): 172-176.
- Beil, R., and G. Delemeester. "The Double Oral Auction: Is It an Effective Teaching Tool?" In *Proceedings, Economics and the Classroom Conference* (pp. 12-32). Idaho State University and Prentice-Hall Publishing Co., March, 1999.
- Bloom, B.S., and D.R. Krathwohl. *Taxonomy of Educational Objectives: The Classification of Educational Goals, by a committee of college and university examiners. Handbook I: Cognitive Domain*. New York: Longmans, Green, 1956.

- Brauer, J., and G. Delemeester. "Games Economists Play: A Survey of Non-Computerized Classroom-Games for College Economics." *J. Econ. Surveys* 15(April 2001):221-236.
- Cardell, N.S., R. Fort, W. Joerding, F. Inaba, D. Lamoreaux, R. Rosenman, E. Stromsdorfer, and R. Bartlett. "Laboratory-Based Experimental and Demonstration Initiatives in Teaching Undergraduate Economics." *Amer. Econ. Rev.* 86(May 1996):454-459.
- Dickie, M. "Do Classroom Experiments Increase Learning in Introductory Microeconomics?" *J. Econ. Educ.* (Summer 2006):267-288.
- Fels, R. "This is What I Do, and I Like It." *J. Econ. Educ.* 24(Fall 1993):365-70.
- Fraas, J.W. "The Use of Seven Simulation Games in a College Economics Course." *J. Experimental Educ.* 48(Summer, 1980):264-280.
- Frank, B. "The Impact of Classroom Experiments on the Learning of Economics: An Empirical Investigation." *Econ. Inquiry* 35(October, 1997):763-769.
- Gremmen, H., and J. Potters. "Assessing the Efficacy of Gaming in Economic Education." *J. Econ. Educ.* (Fall, 1997):291-303.
- Hansen, W.L. "What Knowledge is Most Worth Knowing – For Economics Majors?" *Amer. Econ. Rev.* 76(May 1986):149-152.
- Hauptert, M.J. "An Experiment in Comparative Advantage." *J. Econ. Educ.* 27(winter 1996):37-44.
- Herz, B., and W. Merz. "Experiential Learning and the Effectiveness of Economic Simulation Games." *Simulation and Gaming* 29(June 1998):238-250.
- King, H. "A Neophyte's Cost-Benefit Analysis of Classroom Experiments and Simulations in Introductory Economics." Discussion Paper #79, Department of Economics, University of Regina, July, 1999.

- Kolb, D.A. "Learning Styles and Disciplinary Differences." *The Modern American College*. A. W. Chickering and Associates, eds., pp. 232-255. San Francisco: Jossey-Bass Inc., 1981.
- Krathwohl, D.R., B.S. Bloom, and B.B. Masia. *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook II: Affective Domain*. New York: David McKay Co., Inc., 1964.
- Mullin, D., and G. Sohan. "Benefit Assessment of Classroom Experimental Economics." Working Paper, United States Air Force Academy, April, 1999.
- Nelson, R.G., and R.O. Beil, Jr. "Pricing Strategy under Monopoly Conditions: An Experiment for the Classroom." *J. Agr. and Applied Econ.* 26(July 1994):287-298.
- Noussair, C., and J. Walker. "Student Decision Making as Active Learning: Experimental Economics in the Classroom." *Teaching Economics to Undergraduates: Alternatives to Chalk and Talk*, W.E. Becker and M. Watts, eds., pp. 49-77. Northampton, MA: Edward Elgar. 1998.
- Smith, V.L. "Experimental Methods in Economics." *The New Palgrave: Allocation, Information, and Markets*. J. Eatwell, M. Milgate, and P. Newman, eds., pp. 94-111. New York: W.W. Norton and Company, 1989.
- Wentworth, D.R., and D.R. Lewis. "A Review of Research on Instructional Games and Simulations in Social Studies Education." *Social Education* 37(May, 1973):432-440.
- Yandell, D. "Effects of Integration and Classroom Experiments on Student Learning and Satisfaction." In Proceedings, *Economics and the Classroom Conference*, 1999. pp. 4-11. Idaho State University and Prentice-Hall Publishing Co.