

On Academics

TEACHING RESEARCH METHODS TO GRADUATE STUDENTS IN PUBLIC HEALTH

MELBOURNE F. HOVELL, PhD, MPH

MARC A. ADAMS, MPH

GEORGE SEMB, PhD

In *The Future of Public Health*, the Institute of Medicine (IOM) recommended improved instruction in quantitative methods and research skills in public health.¹ The IOM emphasized the ethical commitment of public health professionals to use quantitative knowledge to reduce suffering and enhance quality of life. A Harvard study found that two-thirds of the alumni ranked quantitative skills as directly applicable to their profession, supporting IOM's position.² The IOM report recognized inadequate education and training of the workforce as one of the causes of disarray.¹

The Public Health Faculty/Agency Forum addressed the educational dimensions of IOM's findings, and core competencies were developed for Master of Public Health (MPH) students.³ Embedded in these competencies were specific objectives: the ability to (1) define, assess, and understand factors that lead to health promotion and disease prevention, (2) understand research designs and methods, (3) make inferences based on data, (4) collect and summarize data, (5) develop methods to evaluate programs, (6) evaluate the integrity of data, and (7) present scientific information to professional and lay audiences. The Forum also suggested that faculty review, evaluate, and refine established courses and develop new ones. However, the Forum did not outline teaching strategies to achieve these objectives.⁴ Employing improved teaching methods implies stronger evaluations of existing classes from which stronger teaching methods may be derived.

Two methods of instruction have been empirically validated: Keller's Personalized System of Instruction (PSI)⁵ and Socratic-Type Programming.^{5,6} Keller emphasized the following components: (1) identification of specific terminal skills or knowledge, (2) individualized instruction, (3) use of teaching assistants, and (4) use of lectures as a way to motivate students, not just to transmit information. PSI has been more effective than conventional instruction in a variety of educational settings and has increased student

achievement and consumer satisfaction more than conventional instruction.^{7,8}

Socratic-Type instruction emphasizes student responses during lecture, rather than presentations by instructors.^{6,9,10} PSI calls for active responding in the classroom and the Socratic Method is one means of achieving this response. Moran and Malott summarize the empirical evidence and most reliable teaching technologies to date, from which it is clear that frequent quizzing coupled with Socratic procedures contributes importantly to learning.¹¹ A recent study of active learning assessed the use of study guide questions for pending lectures, and outcomes were greater for the study guide than for traditional lecturing.¹²

Research methods should be taught and evaluated by the very methods being taught. This requires experimental or quasi-experimental methods to model the application of science for teaching evaluations. Connelly and associates called for the development of reliable and valid measurement of specified competencies in public health education.¹³ Sarrel found that the use of pretests, interim testing, and a formal posttest evaluation of their training program demonstrated clear learning and strengthened their program.¹⁴

The public health infrastructure relies on people who have the skills to deliver and evaluate interventions. Common teaching methods, such as lectures, limited homework assignments, and final exams, may not maximize learning and do not enable assessment of learning attributable to specific teaching methods. Thus, research methods should be applied to evaluate public health teaching methods to refine them continuously over time. We propose an educational model to assess behavior change over time and describe a course that effectively teaches principles of research design using PSI and Socratic Methods.

METHODS

Participants

Graduate students' performance in Public Health 607 (Research Methods and Proposal Writing) was assessed over an 11-year period (1992 through 2002). Approximately 62% of each class was female, with class sizes varying from 16 to 35 students. Twenty-five out of the 303 students were in the process of or had already completed a doctoral degree. Table 1 shows demographics and student academic backgrounds.

Procedures

Data collection and design. The class was taught once a year with a pretest/posttest design replicated across

Table 1. Class size, gender makeup, GPA, GRE, pretest, and posttest scores over 11 years

Year	Class size	Gender (percent female)	N	GPA M (SD)	N	GRE verbal M (SD)	N	GRE quantitative M (SD)	N	Pretest M (SD)	N	Posttest M (SD)	N
1992	25	88	25	3.0 (0.40)	15	480 (70)	14	538 (78)	14	21.08 (7.98)	25	78.15 (17.13)	24
1993	35	74	35	2.9 (0.33)	29	508 (73)	27	530 (101)	27	12.21 (5.60)	35	80.00 (8.57)	35
1994	27	75	24	3.2 (0.40)	24	485 (79)	21	562 (115)	21	9.83 (4.42)	27	84.81 (7.89)	27
1995	28	78	27	3.0 (0.42)	23	494 (110)	23	523 (114)	23	13.21 (5.24)	28	85.61 (7.76)	28
1996	16	62	13	3.0 (0.32)	13	470 (61)	12	557 (103)	12	17.50 (10.23)	16	82.75 (10.76)	16
1997	25	79	24	3.2 (0.39)	18	486 (82)	16	527 (120)	16	14.86 (6.78)	25	89.30 (7.69)	23
1998	31	81	26	3.4 (0.37)	27	495 (105)	22	553 (94)	22	21.31 (8.49)	31	83.15 (12.59)	31
1999	31	80	30	3.5 (0.34)	26	485 (85)	23	564 (91)	23	17.74 (5.84)	31	89.44 (6.77)	29
2000	25	81	21	3.4 (0.34)	22	497 (88)	20	584 (94)	20	14.38 (10.69)	25	85.50 (8.86)	25
2001	32	78	32	3.3 (0.37)	30	481 (96)	27	559 (130)	26	21.02 (8.19)	32	78.25 (9.96)	32
2002	28	82	28	3.4 (0.39)	21	426 (109)	17	529 (125)	16	16.48 (9.35)	28	79.43 (9.22)	28
Mean, total, or percent	303	78	285	3.2 (0.41)	248	485 (90)	222	548 (107)	220	16.32 (8.37)	303	83.14 (10.59)	298

GPA = grade point average

GRE = graduate record examination

SD = standard deviation

M = mean

the 11 years. Students entering the MPH program came from different fields with at least a bachelor's degree, but many held master's or doctoral degrees. A comprehensive pretest consisting of short-answer and essay items was administered on the first day of class to measure students' prior knowledge of research design. Questions covered content that ranged from concepts of measurement and sources of variance, to operational means of obtaining associations, to nonexperimental and experimental designs. The questions measured knowledge of research designs and related concepts, proper use of designs to applied circumstances, and ability to critique research methodology and develop novel studies. The Figure shows example questions. Variations of the same questions were included in a comprehensive final exam.

To encourage high effort on the pretest, students were told that anyone who scored 80% or higher would be exempt from quizzes and that a one-letter grade bonus would be assigned to their final grade in the course. No student met the 80% criterion. Graduate assistants and the instructor graded the tests with discrepancies judged by the instructor. Students' gender, prior degrees, graduate record examination (GRE), and grade point average (GPA) were used to predict change in examination performance.

Students also evaluated 13 aspects of the class and instructor at the end of every semester. These evaluations were administered by volunteer students. The class instructor was not present during the evaluations and was blind to student ratings. Completed evaluations were given to the graduate school of public health administrator, who produced aggregated summary scores the following semester. Evaluation questions are provided in Table 2. Students responded to each item on a five-point ordinal scale, with anchors of 1 = strongly disagree to 5 = strongly agree.

Class content. The class emphasized Hill's postulates of causal inference and their foundation for research.¹⁵

Designs taught included fully controlled trials, quasi-experimental, within subject time series, case control, cross-sectional, cohort, and nested case control. Each design was evaluated on its strengths and weaknesses for concluding causal associations, including strength of association, consistency of association, temporal order, ruling out alternative explanations, and theoretical plausibility. Concepts such as variance, reliability, validity, fidelity, bias, and confounding were reviewed in the context of each design.

Multicomponent intervention. The class met for two hours and 40 minutes every week for 15 weeks. Six short-answer/short-essay quizzes were administered every other week. Quiz questions required students to generalize from major course concepts and provide novel examples to demonstrate understanding. The questions were similar in format and content to questions on the pretest and final exam. Students who scored less than 80% to 85% (the criterion varied from semester to semester) on a quiz had an option to take a second form of the quiz the following week. Overall, quizzes accounted for 40% of each student's grade in the course.

Lecture and lab time were divided evenly. Lecture was an interactive period where the instructor, using the Socratic Method, asked the students increasingly demanding questions. Correct answers were praised, while incorrect answers prompted the instructor to probe for more basic understanding while avoiding explicit criticism. The main topics of discussion were concepts presented in the readings. Lab time provided clarification of concepts by the teaching assistant in a more relaxed environment. Students were able to practice using the concepts and ask questions in smaller groups or a more casual environment. Exercises were used to provide research experience. Guest speakers offered specialized information such as library research methods and ethics in research.

Over years of administration, two to three texts were

Figure. Examples of questions presented on pretests and posttests

1. Define what is meant by an association between two variables, and give a NOVEL example of both a positive and a negative association.
2. Using a NOVEL example, explain how the following designs meet some or all of Hill's postulates.

(a) Case control	(e) Multiple baseline
(b) Cross-sectional	(f) Reversal design
(c) Cohort	(g) Randomized controlled trial
(d) Quasi-experimental	
3. Explain why many practitioners do not believe results from research and do not adopt new treatment or prevention procedures based on research results using the concepts of variance and ecological fallacy.
4. Define and give a NOVEL example of (a) reliability and (b) validity.

Table 2. Students' aggregated quantitative evaluations from 1997 to 2002

Question ^a	Median score	Median range	
		Minimum	Maximum
Q1 Course objectives and procedures for evaluation were made clear at the beginning of the semester.	4.25	4.0	5.0
Q2 Content, assignments, activities, and evaluation procedures were appropriate for course objectives.	4.00	4.0	5.0
Q3 Content (lectures, handouts, reading assignments) included current, state-of-the-art information.	4.00	4.0	4.0
Q4 Course resources (text, handouts, guests, slides, film) were useful.	4.00	3.0	4.0
Q5 Planned course content was reasonably covered by the end of the semester.	4.00	4.0	4.0
Q6 Overall, the course was personally rewarding.	4.00	4.0	5.0
Q7 Instructor was knowledgeable on the subject matter.	5.00	4.5	5.0
Q8 Instructor was prepared to lead class activities.	5.00	4.0	5.0
Q9 Instructor was professional and understanding in relationships with students.	4.50	3.0	5.0
Q10 Instructor communicated effectively in class and/or conferences.	4.00	3.0	5.0
Q11 Instructor encouraged interaction (i.e., questions, expression of opinion).	5.00	4.0	5.0
Q12 Instructor provided adequate feedback regarding student performance in course.	3.75	3.0	4.0
Q13 Overall, this instructor was an effective teacher and facilitated learning the material.	4.00	4.0	5.0
Total	4.00		

^aResponses were based on a five-point scale: 1 = strongly disagree, 2 = disagree, 3 = neither agree nor disagree, 4 = agree, 5 = strongly agree.

used. The primary text was a current edition of *Designing Clinical Research*.¹⁶ The second text emphasized quasi-experimental and time-series designs.¹⁷ These were supplemented by research articles to illustrate issues and methodological errors.

In addition to answering specific questions concerning each reading, homework assignments required incremental preparation of components of a grant application based on the PH398 form used by the National Institutes of Health. These included specific aims, operational definitions, literature review, research design, subjects, procedures, and protection of human subjects. Each student was also required to prepare a written critique of a published article. Critical feedback was provided on writing style, use of concepts, design, and understanding of methodological issues. These assignments were iterative and cumulative, leading to both a draft research proposal and refined article critique by the end of the semester. Homework was worth 15% of the semester grade.

Statistical analyses. Descriptive statistics such as percents, means, and medians were computed for each class. Separate analyses for classes and students as the unit of analyses were conducted. An ANOVA model was used to evaluate whether classes differed by year on pretest and posttest scores, GPA, and GRE scores. Student change scores were computed by subtracting pretest from posttest scores, with mean differences evaluated by a paired sample *t*-test. A Pearson correlation coefficient determined associations among students' covariates, test scores, and change scores. We estimated a regression model with students' pre- to

posttest change as the dependent variable and pretest, gender, class year, doctoral status, GPA, and GRE scores used as covariates. Strength of each predictor was estimated by standardized betas. All analyses were performed using SPSS.¹⁸

RESULTS

Table 1 presents a summary of class size, proportion of female students, GPA, GRE, and pre- and posttest scores over the 11 years. Mean GPA scores ranged per class from 2.9 to 3.5, based on a four-point scale. Mean class GRE verbal scores ranged from 426 to 508, and quantitative scores ranged from 523 to 584. For all 11 years combined, the total mean GPA was 3.2 (standard deviation [SD] = 0.41) and the mean GRE verbal and quantitative scores were 485 (SD=90) and 548 (SD=107), respectively. An ANOVA determined classes differed on mean GPA scores $F(10, 247)=6.03$, $p<0.001$. However, GRE verbal and quantitative scores were not significantly different among years.

Class pretest scores were consistently low and ranged from 9.8% to 21.3% correct, while posttest scores were consistently higher, ranging from 78.2% to 89.4% correct. The mean pre- and posttest scores were 16.3% (SD=8.4) and 83.1% (SD=10.6) correct, respectively. With limited variance around the posttest scores, a high proportion of students attained more than 75% correct by the end of the semester. Pre- to posttest change ranged from 57.1% to 75.0%. The 11-year mean increase from pre- to posttest for all students was 66.8%. This difference was significant $t(297)=98.72$, $p<0.001$ as expected using a paired sample

t-test. For comparison, students pursuing or holding a doctoral degree had comparable pretest ($M=23.26\%$, $SD=7.96$), posttest ($M=86.56\%$, $SD=10.52$), and change ($M=63.30\%$, $SD=10.52$) scores.

To estimate the likelihood that consistent change in scores from pre- to posttest was attributable to the class methods, a binomial probability with a 50% chance of changing at least 50 percentage points was computed for 11 consecutive years of change. The probability of obtaining the observed result was less than 0.001.

A Pearson correlation coefficient was used to determine associations among students' characteristics, test scores, and change scores. There was a significant negative correlation between students' class year and pretest scores ($r=-0.19$, $p=0.003$), but no significant correlation between class year and posttest scores ($r=-0.02$, $p=0.78$). Significant positive relationships were found between students' pretest scores and GRE verbal ($r=0.31$, $p<0.001$), GRE quantitative ($r=0.25$, $p<0.001$), and GPA ($r=0.37$, $p<0.001$) scores. Similar significant positive associations were seen with students' posttest scores and GRE verbal ($r=0.26$, $p<0.001$), GRE quantitative ($r=0.30$, $p<0.001$), and GPA ($r=0.34$, $p<0.001$) scores. While GRE quantitative scores had a marginally significant positive correlation with the pre- to posttest change scores ($r=0.11$, $p=0.09$), GRE verbal and GPA scores were not associated with the students' change scores ($r=0.03$, $p=0.64$ and $r=0.05$, $p=0.45$). Pretest, GPA, and GRE scores were included in further multivariate analyses.

Table 3 presents students' change scores regressed on predictor variables. The model explained 29.4% of the variance. Variables significantly associated with change scores included students' pretest scores (stan-

dardized beta = -0.53), GPA (standardized beta = 0.24), GRE quantitative score (standardized beta = 0.19), and class year (standardized beta = -0.15). The results indicate that, for every one standard deviation decrease from the mean pretest score, posttest scores increased 0.53 SDs, after controlling for all other variables. This suggests that students with poorer understanding of the material at pretest may benefit most from the class. Moreover, students with higher GPA and GRE quantitative scores showed greater change.

Students ranked qualitative aspects of the class each year. However, data were available only from 1997 to 2002 (Table 2). Overall, students ranked the class favorably. Median scores ranged from 3.0 to 5.0, with the total median score over the six years equaling 4.0, or "agree," with the 13 positively framed evaluation questions.

DISCUSSION

This article describes an effective way to teach research design to graduate students in public health. The teaching method specified objectives, frequent assessments of student behavior, individualized feedback, and the opportunity to do assignments until mastered. Pre- to posttest change demonstrated dramatic increases in competencies by semester's end. These changes were replicated over 11 consecutive cohorts.

On the pretest, most students demonstrated a lack of conceptual understanding of public health science. Many students demonstrated what might best be described as a memorized understanding of procedures with little understanding of how specific procedures controlled for error, or without the ability to use novel

Table 3. Linear regression model for students' change scores^a

Variables ^b	<i>b</i>	SE	β	95% CI	P-value
Pretest score	-0.770	0.10	-0.53	-0.58, -0.96	<0.001
GPA	6.860	1.98	0.24	2.95, 10.77	0.001
GRE quantitative	0.021	0.01	0.19	0.01, 0.04	0.008
Year	-0.520	0.24	-0.15	-0.04, -0.99	0.033
Doctoral degree	10.980	9.86	0.07	-8.46, 30.43	0.267
GRE verbal	0.005	0.01	0.04	-0.01, 0.02	0.585
Gender	-0.890	1.91	-0.03	-4.65, 2.87	0.640

^a $R^2 = 0.294$

^bVariables are ordered according to beta weight, from highest to lowest.

b = unstandardized coefficient

SE = standard error

β = standardized coefficient

CI = confidence interval

GPA = grade point average

GRE = graduate record examination

methods to solve research problems. While this class did not equip students with advanced research ability, it moved them in that direction through active participation and feedback. The pretest also showed the need for more training of most behavioral science majors in public health, including an important proportion of students who had completed relatively advanced training in medicine. This feedback tended to have a humbling effect on students that motivated them to participate fully in the class.

The class was based on Keller's Personalized System of Instruction and the Socratic methods. Teaching assistant tutorials, repeated quizzes, and the Socratic methods resulted in reliable mastery of the content. Most students ended the class with a B grade or higher. The few who obtained a C were invited to repeat the class and all obtained a B or higher with replication. Moreover, those who obtained a C grade demonstrated remarkable change from about 10% correct answers on pretests to more than 70% correct at posttest.

The bivariate analyses showed weak relationships between most demographic and baseline knowledge characteristics and final change from pretest. Scores at pretest and posttest performance tended to be weakly related to pre- to posttest change, but this is essentially a mechanical relationship as they are components of the change score. Examination of GRE and GPA scores showed weak prediction of change in bivariate analyses. This suggests that they did not predict change in the context of this class.

The multivariate analysis showed that pretest scores, GPA, quantitative GRE scores, and year of enrollment were significant predictors of change scores. However, pretest scores are again a function of the mechanical relationship as a component of the change score, providing more opportunity for change with lower-level pretest scores. GPA and GRE predictors are face valid and probably reflect past academic skills, both specific knowledge such as quantitative skills, and general knowledge of how to study for classes. Year of enrollment may reflect subtle changes in the demography of the students or in the teaching procedures. However, all of these factors explained less than 30% of the variance in the pre- to posttest performance changes. This provides circumstantial evidence that the changes achieved were due to the teaching procedures. This evidence is strengthened when one examines the role of doctoral training (mostly physicians), verbal GRE, and gender, none of which reached significance as predictors of change. These findings suggest that individuals with a relatively high GPA and relatively advanced degrees, including a doctoral degree, can benefit from a teaching technology of the sort tested here.

Limitations

This study lacked random assignment to a comparison group for more complete control for alternate explanations of change. However, the class included interactive feedback through biweekly quizzes, frequent student and instructor interaction, homework, and tutoring. These repeated interactions served as student performance "process measures" that corroborated the pre- to posttest changes observed. Moreover, our results showed that the likelihood of obtaining 11 consecutive years of substantial change was less than one in 1,000, with students differing in academic achievement and backgrounds. Although the efficacy of PSI and the Socratic Method were not compared to other pedagogical methods, the current research provides substantive evidence for attributing change in performance to the current instructional methods, and suggests a high level of generalizability.

Standards for teaching public health

The IOM called for more quantitative instruction in public health. This article provides a model for well-researched PSI and related Socratic teaching methods that can be applied to all graduate public health instruction.¹⁹ While more formal control procedures could be used in the experimental evaluation of similar classes, we recommend that, at a minimum, objective measures of knowledge be assessed routinely on a pretest and posttest basis. Doing so provides quasi-experimental evidence of learning attributable to the specific class, and it provides a model to students for the routine use of research methods in the delivery of services. These procedures also lend to use by academic personnel committees for remedial assistance for instructors in those instances where outcomes do not appear satisfactory. Such procedures should supplement subjective student evaluations. The Association of Schools of Public Health and the Council on Education in Public Health might consider criteria for evaluating public health instruction that follows this model. Such policies and teaching procedures could meet the standards recommended by the IOM.

Melbourne Hovell is a Professor at the Center for Behavioral Epidemiology and Community Health (CBEACH) at San Diego State University Graduate School of Public Health in San Diego, CA. Marc Adams is a doctoral student at CBEACH. George Semb is a Professor in the Department of Applied Behavioral Sciences at the University of Kansas in Lawrence, KS.

Research for this article was conducted with intramural support from CBEACH.

Address correspondence to: Melbourne F. Hovell, PhD, MPH, Center for Behavioral Epidemiology and Community Health, Graduate School of Public Health, San Diego State University,

9245 Sky Park Ct., Ste 230, San Diego, CA 92123; tel. 858-505-4770, ext. 103; fax 858-505-8614; e-mail <mhovell@projects.sdsu.edu>.

REFERENCES

1. Committee for the Study of the Future of Public Health, Division of Health Care Services, Institute of Medicine. *The future of public health*. Washington: National Academy Press; 1988. Also available from: URL: http://www.nap.edu/catalog.php?record_id=1091 [cited 2007 Apr 23].
2. Kahn K, Tollman SM. Planning professional education at schools of public health. *Am J Public Health* 1992;82:1653-7.
3. Sorensen AA, Bialek RG. The public health faculty/agency forum: linking graduate education and practice: final report. Atlanta: Bureau of Health Professions, Health Resources and Services Administration and Public Health Practice Program Office, Centers for Disease Control and Prevention (US); 1991.
4. Potter MA, Pistella CL, Fertman CI, Dato VM. Needs assessment and a model agenda for training the public health workforce. *Am J Public Health* 2000;90:1294-6.
5. Keller FS. "Good-bye teacher . . ." *J Appl Behav Anal* 1968;1:79-89.
6. Mahan HC. The use of Socratic type programmed instruction in college courses in psychology. Paper read at Western Psychological Association; May 1967; San Francisco.
7. Kulik C-LC, Kulik JA, Cohen PA. Instructional technology and college teaching. *Teaching of Psychology* 1980;7:199-205.
8. Kulik JA, Kulik C-LC, Cohen PA. A meta-analysis of outcome studies of Keller's Personalized System of Instruction. *Am Psychol* 1979;34:307-18.
9. *The American Heritage Dictionary of the English Language*. 4th ed. Boston: Houghton Mifflin; 2006.
10. Binker A. Socratic questioning. In: Paul RW, editor. *Critical thinking: what every person needs to survive in a rapidly changing world*. Dillon Beach (CA): Foundation for Critical Thinking; 1993. p. 360-90.
11. Moran DJ, Malot RW, editors. *Evidence-based educational methods*. San Diego: Elsevier Academic Press; 2004.
12. Saville BK, Zinn TE, Neef NA, Van Norman R, Ferreri SJ. A comparison of interteaching and lecture in the college classroom. *J Appl Behav Anal* 2006;39:49-61.
13. Connelly J, Knight T, Cunningham C, Duggan M, McClenahan J. Rethinking public health: new training for new times. *J Manag Med* 1999;13:210-7.
14. Sarrel PM, Sarrel LJ, Faraclas WG. Evaluation of a continuing education program in sex therapy. *Am J Public Health* 1982;72:839-43.
15. Hill A. The environment and disease: association or causation? *Proc R Soc Med* 1965;58:295-300.
16. Hulley SB, Cummings SR. *Designing clinical research: an epidemiologic approach*, 1st ed. Baltimore: Williams & Wilkins; 1988.
17. Hayes SC, Barlow DH, Nelson-Gray RO. *The scientist practitioner: research and accountability in the age of managed care*. Boston: Allyn & Bacon; 1999.
18. SPSS, Inc. *SPSS for Windows: Version 12.0.1*. Chicago: SPSS, Inc.; 2004.
19. Greenhoot AF, Semb G, Colombo J, Schreiber T. Prior beliefs and methodological concepts in scientific reasoning. *J Appl Cogn Psychol* 2004;18:203-21.