

**Using the Post-then Method to Assess Learner Change  
Presentation at the AAHE Assessment Conference  
June 15, 2000, Charlotte, North Carolina**

Karl Umble<sup>1</sup>, Ph.D., M.P.H. ([kumble@sph.unc.edu](mailto:kumble@sph.unc.edu)), Vaughn Upshaw<sup>2</sup>, Dr.P.H., Stephen Orton<sup>1</sup>, Ph.D., Kelly Matthews<sup>2</sup>, Ph.D. (candidate)

<sup>1</sup>The North Carolina Institute for Public Health and <sup>2</sup>Department of Health Policy and Administration, School of Public Health, The University of North Carolina at Chapel Hill

- I. Self-reported Measurements in Educational Research and the Pre-Post Design
  - A. Self-reported measurements are frequently used in educational settings (tests, surveys, interviews)
  - B. Response-shift bias weakens the validity of traditional pre-post design for self-reported measurements.

Examples: measurements of assertiveness, or skill in interviewing, pre and post. Internal scale of participants changed during the training, so that the pre-test and the post-test are actually using different measurement scales. The shift of the internal scale results in the “response shift bias” in the results. Tendency for pre-ratings to be elevated, leading to findings of negative or reduced or non-significant treatment effects using this design (Howard, 1979).

- II. The Post-then method
  - A. Description and Advantages
    - 1. Also known as the “retrospective pre-test” or the “post-then-pre” design
    - 2. Eliminates response-shift bias, often finds significant changes in situations where a pre-post design does not.
    - 3. Independent measures of changes in constructs (such as student ratings of teachers) usually correlate more with post-then differences than pre-post differences, leading to beliefs that these methods are more valid (Howard, 1979, Skeff et al., 1992). Also, interviews with learners often reveal response-shift.

Example:

Item: Rate your skill in qualitative data collection through interviewing on a scale of 1 (not at all skilled) to 5 (very skilled).

Pre-Post Design

Pre-rating: 4

Post-rating: 4

Post-then Design

Retrospective pre-test rating: 2

Post-rating: 4

B. Examples

**TABLE 1** – Representative Studies Using the Post-Then Method

<b>Study</b>	<b>Knowledge</b>	<b>Skills</b>	<b>Self- efficacy</b>	<b>Attitudes</b>	<b>Behavior</b>
Howard, Ralph, Gulanick, et al.(1979)		Helping skills - students		Dogmatism – Air Force personnel	Assertiveness – Air Force women
Howard & Dailey (1979)		Interviewing – HR professionals			
Howard, Schmeck, & Bray (1979)	College students' knowledge in course				
Howard, Dailey, & Gulanick (1979)		Interpersonal effectiveness			
Bray & Howard (1979)		Teaching			
Howard, Millham, Slaten & O'Donnell (1980)					Assertiveness
Pohl (1982)	College students' statistics knowledge				
Zwiebel (1987)				Acceptance and empathy towards disabled	
Rhodes & Jason (1987)					Tobacco and drug use, teens
Rockwell & Kohn (1989)					Nutritional
Levinson, Gordon, & Skeff (1990)					Teaching, medical school faculty
Skeff, Stratos, & Bergen (1992)					Teaching, medical school faculty
Donehoo (1998)	Knowledge, nursing students				
Umble, Polhamus, & Farel (2000)					Data collection and analysis, health prof
Umble, Sollecito, Endress (2000)			Public health leadership skills		
Umble, Upshaw, Orton, Matthews (2000)					Public health management skills

### III. Example Instruments

#### A. Self-efficacy

For measuring gains in self-efficacy (Umble, Sollecito, & Endress, 2000). This appeared on a final survey (web-based) for participants in a distance learning MPH degree program for working public health professionals. There were 14 self-efficacy items corresponding with skills the program was intended to teach. For more on self-efficacy, see Maibach & Murphy (1995) and Applebaum (1996).

#### *Example Items, Analysis, and Findings*

Each question below describes a situation in which you are being asked to complete a task. It then asks you to **rate your confidence that you could have performed that task on August 1, 1997, just before you entered the MPH program.** Then it asks you to rate your confidence that you could perform these same tasks **today**. For this section, use the following rating scale:

- |                            |                          |                        |                            |
|----------------------------|--------------------------|------------------------|----------------------------|
| A. Not at all<br>confident | B. Somewhat<br>confident | C. Mostly<br>confident | D. Completely<br>confident |
|----------------------------|--------------------------|------------------------|----------------------------|

#### *Analysis:*

Paired samples t-test results indicate a significant difference between the retrospective pre-test and the post-test for all items ( $p < .001$ ), and that self-efficacy scores on post-test were almost double those on retrospective pre-test.

Question	Rate your confidence that you could have performed the task <b>on August 1, 1997<sup>a</sup></b>	Rate your confidence that you could perform the task <b>today<sup>a</sup></b>	P
Your supervisor asks you to conduct a community health needs assessment. You will be expected to collect, analyze, and interpret information on the health status, health needs, and health services available in your county. How confident are you that you could conduct this health needs assessment?	1.4	3.6	<.001
Your supervisor asks you to do a literature search on a health problem facing your community. You conduct the search and find twelve important original scientific studies of methods to address the problem. How confident are you that you could critically appraise the articles for their scientific merit and for the validity of their conclusions?	1.7	3.6	<.001

## B. Skills

Farel, Polhamus, & Umble (2000) used the retrospective pre-test method to measure perceived skill changes of maternal and child health staff in state and local public health departments who had completed a 1 year, 6 module Web-based training course in data collection and analysis.

### *Example Items*

Each of the following items describes a skill taught in the EDUSIT course. In the BEFORE column, **think back to before the course and rate your skill level before the EDUSIT course started**. In the NOW column, rate your skill level **now**.

Skill	Your Skill Level <b>BEFORE</b> the EDUSIT Course	Your Skill Level <b>NOW</b>
1. Selecting an appropriate secondary data source to investigate a public health problem.	Low 1 2 3 4 5 6 7 High	Low 1 2 3 4 5 6 7 High
2. Conducting a Web search.	Low 1 2 3 4 5 6 7 High	Low 1 2 3 4 5 6 7 High

### *Analysis and Example Findings:*

Question	Skill Level Before (Mean)	Skill Level Now (Mean)	P(T<=t) two-tail <sup>a</sup>
1. Selecting an appropriate secondary data source to investigate a public health problem	3.61	5.78	<.001
2. Conducting a Web search	4.22	6.35	<.001

C. Practices

Farel et al. (2000) also examined practice frequency change for the same skills.

*Example Item*

1. Selecting an appropriate secondary data source to investigate a public health problem.

How many times did you perform this skill in the 6 months **BEFORE** the course?

0 times    1 time    2 or more times

How many times have you performed this skill **SINCE** the course ended?

0 times    1 time    2 or more times

Do you expect to perform this skill in the **FUTURE**?

Yes    No

*Analysis and Example Results*

Analyzing the above items using the Wilcoxin signed rank test (a non-parametric procedure) led to findings of significant changes for practice of most of the skills taught in the course.

<b>Variable Name</b>	<b>Mean Change</b>	<b>Standard Error</b>	<b>P-Value</b>
Selecting an appropriate data source.	0.61	0.17	0.0048
Conducting a Web search.	0.48	0.18	0.0234

## D. Skills

Upshaw, Umble, Orton, & Matthews (2000) used a pretest, a retrospective pre-test, and a post-test design to measure the effects of a 1-year public health management skills training course.

### *Example Item:*

The simple Skills Inventory on this sheet is designed to help you assess your current management skills in several areas the Management Academy have addressed over the past months. We are interested in how your skills have changed during the Management Academy training, and also in acquiring a retrospective account of your skills before you started in the program.

Ask yourself this question with respect to each of the skill areas listed on the survey: Compared to other managers performing similar jobs in comparable organizations, how would you rate your own skills in each of these management areas?

Rank your skill level in each specific area—"motivating staff" and "making goals clear," for instance—on a scale of one to five, with one for "weak skills" through five for "strong skills." In the shaded middle column, rate your skill level before you started the Management Academy. In the last column rate your skill level now.

### **SCALE**

**1** Weak skills   **2** Less than average skills   **3** Average skills  
**4** Better than average skills   **5** Strong skills

Skills	Before MAPH	Now
<i>Managing People</i>		
<i>Motivating staff</i>	1 2 3 4 5	1 2 3 4 5
<i>Making goals clear</i>	1 2 3 4 5	1 2 3 4 5
<i>Selecting staff and orienting them to their roles</i>	1 2 3 4 5	1 2 3 4 5

### *Analysis and Example Results:*

All items were significantly different using the pre-post and the retrospective pre-post ( $p < .0001$ ). Pre-scores and T-values were higher for most items with retrospective design. Below are the average means for some categories of skills on the inventory.

	<b>Pre</b>	<b>Retro_Pre</b>	<b>Now</b>	<b>t-value Pre-Post</b>	<b>t-value Retro-pre</b>
Managing People	3.59	3.27	4.14	9.22	14.69
Managing Finances	2.87	2.84	3.65	10.94	11.90
Managing Projects	3.26	3.13	3.87	6.69	11.95

## Conclusion and Discussion

When self-reported data measuring change are required, the retrospective pre-test method may often be a more valid indicator of program effects than the traditional pre-post method. Certain variables and circumstances may be more appropriate. For example, one training study found that self-reported knowledge gains did not correlate with actual knowledge gain scores (Dixon, 1990). Implications:

1. Use the traditional pre-post design in addition to the retrospective pre-test, to see if a response shift appears to be occurring. Are the results more significant for the retrospective design, and are there differences between the pre-test and the retrospective pre-test scores?
2. The presence of a response shift probably indicates that the retrospective design is more valid (but not necessarily so). It would be a good idea to interview learners to ascertain if they believe a response shift has occurred, or to use another method in addition to the self-report (see Skeff et al., 1992).
3. It is a good idea to use multiple methods to assess change, such as behavioral observations, asking co-workers or supervisors for ratings, role plays, and interviews. However, do not assume that any of these methods are necessarily more valid than the self-report (Howard, 1990). Each method will have advantages and problems in every circumstance.
4. It may also be combined with alternative assessment methods such as journals that ask participants to reflect on how a program has helped them. Reflection on gains made and significance of those gains for one's theory of practice and actual practice might enable the learner to actually gain more from the intervention.

## References

- Aiken, L.S., & West, S.G. (1990). Invalidity of true experiments: Self-report pretest biases. Evaluation Review, 14, 374-390.
- Applebaum, S.H. (1996). Self-efficacy as a mediator of goal setting and performance: Some human resource applications. Journal of Managerial Psychology, 11(3), 33-48.
- Bray, J.H., Maxwell, S.E., & Howard, G.S. (1984). Methods of analysis with response-shift bias. Educational and Psychological Measurement, 44(4), 781-804.
- Dixon, N.M. (1990). The relationship between trainee responses on participant reaction forms and posttest scores. Human Resource Development Quarterly, 1(2), 129-137.
- Donehoo, L.P. (1998). The relationship of response-shift bias to the self-assessment of learning achievement in practical nurse programs in postsecondary vocational education. Unpublished doctoral dissertation, Department of Adult Education, The University of Georgia, Athens.
- Farel, A., Polhamus, B. & Umble, K. Unpublished data. Department of Maternal and Child Health, School of Public Health, University of North Carolina at Chapel Hill.
- Howard, G.S. (1980). Response shift bias. A problem in evaluating interventions with pre/post self-reports. Evaluation Review, 4(1), 93-106.
- Howard, G.S. (1981). On validity. Evaluation Review 5(4), 567-576.
- Howard, G.S. (1990). On the construct validity of self-reports: What do the data say? American Psychologist, 45(2), 292-294.
- Howard, G.S., & Dailey, P.R. (1979). Response-shift bias: A source of contamination of self-report measures. Journal of Applied Psychology, 66(2), 144-50.
- Howard, G.S., Dailey, P.R., & Gulanick, N.A. (1979). The feasibility of informed pretests in attenuating response-shift bias. Applied Psychological Measurement, 3, 481-494.
- Howard, G.S., Millham, J., Slaten, S., & O'Donnell, L. (1981). Influence of subject response-style effects on retrospective measures. Applied Psychological Measurement, 5, 144-150.
- Howard, G.S., Ralph, K.M., Gulanick, N.A., Maxwell, S.E., Nance, S.W., & Gerber, S.K. (1979). Internal invalidity in pre-test-post-test self-report evaluations and a re-evaluation of retrospective pre-tests. Applied Psychological Measurement, 3, 1-23.
- Howard, G.S., Schmeck, R.R., & Bray, J.H. J(1979). Internal validity in studies employing self-report instruments: A suggested remedy. Journal of Educational Measurement, 16, 129-135.
- Howard, G.S. (Feb. 1990) On the construct validity of self-reports: What do the data say? American Psychologist, 45(2), 292-294.
- Koele, P. & Hoogstraten, J. (1988). A method for analyzing retrospective pretest/posttest designs: I. Theory. Bulletin of the Psychonomic Society 26(1), 51-54.
- Levinson, W., Gordon, G., & Skeff, K. (1990). Retrospective versus actual pre-course self-assessments. Evaluation & the Health Professions, 13, 445-452.
- Mann, S. (1997). Implications of the response-shift bias for management. Journal of Management Development, 16(5-6), 328-337.

- Mezoff, B. (1981). How to get accurate self-reports of training outcomes. Training and Development Journal, 35(9), 56-61.
- Maibach E, Murphy DA (1995). Self-efficacy in health promotion research and practice: conceptualization and measurement. Health Education Research, Theory, & Practice, 10:37-50.
- Mezoff, B. (1981). Pre-then-post testing: A tool to improve the accuracy of management training program evaluation. Performance and Instruction, 20(8), 10-11.
- Pohl, N.F. (1982). Using retrospective pre-ratings to counteract response-shift confounding. Journal of Experimental Education, 50(4), 211-214.
- Rhodes, J.E. & Jason, L.A. (1987). The retrospective pre-test: An alternative approach in evaluating drug prevention programs. Journal of Drug Education, 17(4), 345-356.
- Rockwell, K. & Kohn, H. (1989). Post-then pre- evaluation. Journal of Extension, 27(2), 19-21.
- Skeff, Stratos, & Bergen (1992). Evaluation & the Health Professions issue from that year, don't have complete reference but it was in there.
- Umble, K.E., Sollecito, W., & Endress, S. (2000). Unpublished data. Program in Public Health Leadership, School of Public Health, University of North Carolina at Chapel Hill.
- Upshaw, V.M., Umble, K.E., Orton, S., & Matthews, K. (2000). Unpublished data. North Carolina Institute for Public Health, School of Public Health, University of North Carolina at Chapel Hill.
- Zweibel, A. (1987). Changing educational counselors' attitudes toward mental retardation: comparison of two measurement techniques. International Journal of Rehabilitation Research, 10(4), 383-389.