



Using a Pretest-Posttest Design to Evaluate Continuing Education Programs

Journal:	Professional Development: The International Journal of Continuing Social Work Education
Article Title:	<i>Using a Pretest-Posttest Design to Evaluate Continuing Education Programs</i>
Author(s):	<i>Thomas R. Barton, Tracy J. Dietz and Linda L. Holloway</i>
Volume and Issue Number:	<i>Vol. 4 No. 1</i>
Manuscript ID:	<i>41032</i>
Page Number:	<i>32</i>
Year:	<i>2001</i>

Professional Development: The International Journal of Continuing Social Work Education is a refereed journal concerned with publishing scholarly and relevant articles on continuing education, professional development, and training in the field of social welfare. The aims of the journal are to advance the science of professional development and continuing social work education, to foster understanding among educators, practitioners, and researchers, and to promote discussion that represents a broad spectrum of interests in the field. The opinions expressed in this journal are solely those of the contributors and do not necessarily reflect the policy positions of The University of Texas at Austin’s School of Social Work or its Center for Social Work Research.

Professional Development: The International Journal of Continuing Social Work Education is published three times a year (Spring, Summer, and Winter) by the Center for Social Work Research at 1 University Station, D3500 Austin, TX 78712. Journal subscriptions are \$110. Our website at www.profdevjournal.org contains additional information regarding submission of publications and subscriptions.

Copyright © by The University of Texas at Austin’s School of Social Work’s Center for Social Work Research. All rights reserved. Printed in the U.S.A.

ISSN: 1097-4911

URL: www.profdevjournal.org

Email: www.profdevjournal.org/contact

Using a Pretest-Posttest Design to Evaluate Continuing Education Programs

Thomas R. Barton, PhD; Tracy J. Dietz, PhD; Linda L. Holloway, PhD

For nearly a decade, all 50 states, the District of Columbia, Puerto Rico, and the Virgin Islands have regulated social work by requiring practitioners to have either a license or certification (American Association of State Social Work Boards, 1996b, p. 29). As of 1996, thirty-five states and the District of Columbia require licensed or certified social workers to participate in continuing education. These states require, on average, 17.5 hours of continuing education per year. The remaining states do not require continuing education to maintain licensure or certification (American Association of State Social Work Boards, 1996a, p. 1).

The ultimate goal of continuing education is to improve practice effectiveness of professional social workers. A crucial first step toward this goal is to increase social workers' knowledge and understanding of effective practice methods, social work ethics and the ethical decision-making process, and the most current and accurate information concerning client populations and emerging social issues. The two main sources of these types of information are professional journals and continuing education programs.

Dietz (1998) discusses three levels of evaluating professional social work continuing education. The first level involves examining the extent to which participants gained cognitive knowledge about the topic. This is often evaluated using a simple pretest-posttest design (See Denning, 1993; Feucht, Stephens, & Gibbs, 1991; and McCowan, McGregor, & LoTempio, 1982). The second level is to assess whether participants leave the training with the ability to practice what they have learned. This can be evaluated by having participants demonstrate their mastery of the practice technique, for example, within the continuing education

setting (See Lindsey, Yarbrough, & Morton, 1987; and Rooney, 1988). The third level concerns whether continuing education participants actually use the training's information in their practice (See Bibus & Rooney, 1995; and Mueller, 1985).

Level-three, and to a somewhat lesser extent, level-two evaluations, can be difficult, expensive, and require a long-term commitment of the evaluators and participants. As the above suggests, however, a number of evaluations have been conducted using these designs. Since the long-term improvement of practice is the ultimate goal of much of continuing professional education, level-three evaluations are the most desirable means of assessing this facet.

The crucial first step toward the goal of improving practice efficacy is to ensure that continuing education participants obtain knowledge about the subject matter. Without this minimum first step, we cannot expect practice to improve. It is therefore important that social workers use an effective design to evaluate the knowledge component of continuing education programs.

This article discusses the methodological issues concerning the simple pretest-posttest design to assess knowledge gain and suggests that despite methodological limitations, it may be useful in conducting level one evaluations of continuing education programs for social workers (Dietz, 1998).

Design Issues

In order to make the case that a change — any change — is the result of some kind of intervention, one must meet the three criteria of cause. First, the researcher must demonstrate that the intervention or cause precedes the effect in time.

Thomas R. Barton, PhD, is an Associate Professor at University of North Texas, Department of Rehabilitation, Social Work, and Addictions.

Tracy J. Dietz, PhD, is an Associate Professor and Chair at Texas Christian University, Department of Social Work.

Linda L. Holloway, PhD, is an Assistant Professor at University of North Texas, Department of Rehabilitation, Social Work, and Addictions.

Correspondence should be addressed to Thomas R. Barton, Department of Rehabilitation, Social Work, and Addictions, P.O. Box 311456, University of North Texas, Denton, TX 76203-1456

Telephone: (940) 565-2023; E-mail: BARTON@UNT.EDU

The authors thank James Humes and Christine Young for their valuable assistance on this project.

Second, there must be an empirical link or correlation between the independent and dependent variables. Third, and most important for this article, "the observed empirical correlation between two variables cannot be explained away as being due to the influence of some third variable that causes both of them" (Rubin & Babbie, 1997, p. 276; See also, Campbell & Stanley, 1963; Cook & Campbell, 1979; Neuman, 1991; Royse, 1995). In regards to social work continuing education evaluation and program evaluation in general, the most difficult of the three criteria for cause to be met is the third, ensuring that a variable other than the intervention or training does not cause changes in participants' knowledge or practice.

The "internal validity" of an evaluation design is established by the combination of these three criteria and the extent to which they are met. In their classic discussion of this topic, Campbell and Stanley (1963) conclude that "internal validity [emphasis in all] is the basic minimum without which any experiment is interpretable: Did in fact the experimental treatments make a difference in this specific experimental instance? ... internal validity is the sine qua non" (p. 5).

Without satisfying these three criteria, continuing education evaluators cannot persuasively make the case that their programs result in changes in participants' knowledge, ethics, or practice skills. Depending upon the research design employed to evaluate the continuing education program, the internal validity can be: relatively strong if a "true" experimental design is used, relatively weak if a "pre"-experimental design is used, and have middle-range strength if a "quasi"-experimental design is used. Each of these broad classes — true, pre, and quasi — contain several specific designs (Campbell & Stanley, 1963).

One of the simplest and most popular designs to use is the one group pretest-posttest, a pre-experimental design with relatively weak internal validity. Utilizing this design, an evaluator of a social work continuing education program administers a ques-

tionnaire to participants prior to the program and administers the same questionnaire to the same participants after the program. The scores on the two sets of questionnaires are then compared to determine if there has been a change in participants' knowledge.

Assuming scores on a knowledge test do indeed increase from the pretest to posttest administration, this design clearly meets the first two criteria for cause: proper time order and empirical link. It is the third criterion, however, that threatens the design's internal validity. In addition to the continuing education session, it is possible that another variable (or variables) caused participants' scores to increase.

The main concern as evaluators is determining if participants' knowledge increases as a result of their participation in a continuing education program. As such, evaluators need to focus on the most relevant threat to internal validity with the simple pretest-posttest design: the problem of "testing [emphasis in original], the effect of the pretest itself. On achievement and intelligence tests, students taking the test for a second time, or taking an alternate form of the test, etc., usually do better than those taking the test for the first time" (Campbell & Stanley, 1963, p. 9). This problem suggests that observed changes in test scores from the pretest to the posttest could be the result of simply taking the pretest, as well as knowledge obtained from the program. Without a control or comparison group, we cannot know how much of the observed change can be attributed to the program and how much can be attributed to testing. This is why the simple pretest-posttest design has low internal validity and is seen as a weak evaluation design by methodologists.

In addition to the problem of testing, the simple pretest-posttest design also suffers from the problems of: (1) history, an event that occurs outside the experiment between the administration of the pretest and the posttest that can affect participants' performance; (2) maturation, a change that occurs within participants as a function of time that can

affect their performance; (3) instrumentation, using one instrument at the pretest and another at posttest; and (4) statistical regression, using the pretest to select either exceptionally high or low performers to continue with the experiment, and then at the posttest observing their scores regress toward the mean (See Campbell & Stanley, 1963; Cook & Campbell, 1979; Neuman, 1991; Royse, 1995; and Rubin & Babbie, 1997).

Because an underlying purpose of continuing social work education is to enhance professionals' knowledge, and because of the simplicity of the pretest-posttest design to evaluate such programs, an experiment was conducted to evaluate the impact testing has on participants' performance on an instrument designed to measure knowledge gain.

Methods

Between December 1996 and July 1998, twelve continuing education training sessions for 236 Community Rehabilitation Programs personnel were conducted to increase their knowledge of how to help clients obtain competitive employment and supported work. Although not specifically targeting social workers, the sessions were open to members of several professions, which is common for professional continuing education. The methodological goal was to design an evaluation for these sessions that would enable the authors to assess the extent to which participants learned the intended material, while simultaneously examining the impact "testing" has on the pretest-posttest design when used to examine knowledge gain.

The trainings were conducted on an "on call" basis. That is, at the beginning of the grant, the evaluators did not know exactly how many sessions would be conducted. Grant personnel informed Community Rehabilitation Programs about the trainings, and relied on them to call and request trainings. Since the evaluators did not know the total number of trainings that would be conducted, and wanted to randomly assign the training groups into Pre Only, Post Only, and Pretest-Posttest

groups, the assignment was made in blocks of three. Prior to any training sessions, the first three forthcoming sessions were randomly assigned to one of the three experimental groups. The evaluators assumed there would be additional training sessions, so additional forthcoming training sessions were assigned to one of the experimental groups in blocks of three. Over the course of the grant period, a total of twelve training sessions were included in the study. One third of the groups were randomly assigned to Pretest Only; one third to Posttest Only; and one third to Pretest-Posttest experimental groups.

All of the trainings were conducted by the same two people using the same techniques (lecture, question and answer, discussion, handouts, etc.), and each session lasted one and three quarter days, for a total of twelve hours. As best as can be determined, between the pretests and posttests, no events (e.g. a news release concerning supported work or an announced change in policy concerning supported work) occurred outside the experiment that could have affected participants between the pre- and posttests, nor did the participants undergo changes due to the passage of time. An identical instrument was used at the pre- and posttests, and no one was excluded from the trainings due to his or her score on the pretest. Hence, in the evaluators' judgment, the most relevant threat to internal validity in the study was testing.

The experiment's design is as follows:

Group 1:	O1	X	
Group 2:	O2	X	O3
Group 3:		X	O4

Each "O" or "observation" indicates when participants completed an identical 15-item instrument designed to obtain their knowledge of competitive employment and supported work (see Appendix for the complete instrument). Each "X" indicates participation in the training session. The pretests were administered prior to the beginning of the trainings, and the posttests were administered just prior to the end.

Using a Pretest-Posttest Design to Evaluate Continuing Education Programs

The evaluation was conducted under two severe constraints. First, the participants came to the training sessions as members of fixed groups. The evaluators were unable to randomly assign participants as individuals to one of the three groups and were consequently forced to assign them as groups. Second, all of the participants received the training. Therefore it was not possible to have a true comparison group that responded to a pretest and a posttest under a similar timeline as members of the other groups but who did not receive the training. However, since the members of Groups One and Two had not participated in the trainings prior to taking the pretest, they effectively serve as a control group.

Hypotheses

The research question posed and the design utilized in this evaluation required the evaluators to test several hypotheses.

Hypothesis One: In order to examine the extent to which the groups were equally matched after random assignment, it was hypothesized that the pretest scores of Groups One and Two would be equal ($O1 = O2$).

Hypotheses Two through Five: A series of hypotheses were posed that one would expect to find if the training increased participants' knowledge of supported work regardless of a testing effect ($O3 > O2$; $O4 > O2$; $O4 > O1$; and $O3 > O1$).

Hypothesis Six: This hypothesis directly addresses the issue of testing. If the testing effect does indeed cause scores to increase because of their second taking of the test, one would expect the Pretest-Posttest Group's posttest scores to be higher than the Posttest Only Group's scores ($O3 > O4$). If, however, testing had no impact on participants' scores, the final hypothesis will be unsupported with the data.

Findings

Table 1 presents selected personal characteristics for all of the 236 participants and is organized by

experimental group. Overall, the participants have a great deal of experience in the rehabilitation field, 7.3 years on average, and 3.1 years at their current jobs. The large standard deviations for both of

Table 1. Participants' Characteristics by Group

	TOTAL (n=236)	GROUPS		
		ONE (Pretest Only) (n = 60)	TWO (Pretest and Posttest) (n = 92)	THREE (Posttest Only) (n = 84)
Total years working in rehabilitation				
Mean	7.3	6.3	7.2	7.9
S.D.	6.1	6.3	6.3	5.9
Years at current job				
Mean	3.1	2.6	3.4	3.3
S.D.	3.2	3.0	3.4	3.1
Gender				
Male	34.8%	34.5%	29.7%	40.5%
Female	65.2%	65.5%	70.3%	59.5%
Age				
Mean	37.5	37.7	37.4	37.4
S.D.	9.3	9.4	10.0	8.7
Race				
African American	22.6%	50.0%	15.3%	12.2%
Caucasian	51.1%	25.9%	71.8%	46.3%
Asian	0.9%	1.9%	1.2%	0.0%
American Indian	0.9%	1.9%	1.2%	0.0%
Hispanic	23.1%	20.4%	8.2%	40.2%
Other	1.4%	0.0%	2.4%	1.2%
Do you have a disability?				
Yes	7.6%	5.2%	11.6%	4.9%
No	92.4%	94.8%	88.4%	95.1%
Education				
Some high school	0.8%	0.0%	1.1%	1.2%
High school graduate	15.7%	18.3%	18.5%	10.7%
Some college	44.1%	43.3%	34.8%	54.8%
College graduate	22.9%	15.0%	26.1%	25.0%
Some graduate work	8.5%	11.7%	7.6%	7.1%
Graduate degree	8.1%	11.7%	12.0%	1.2%
Licenses or Certificates Held				
None	58.9%	77.8%	58.2%	68.0%
Rehabilitation	4.2%	3.7%	5.1%	5.3%
Nursing	3.4%	0.0%	5.1%	5.3%
Education	5.1%	3.7%	7.6%	5.3%
Social Work	3.8%	5.6%	2.5%	5.3%
Other	12.7%	9.3%	21.5%	10.7%
Have you received other training in supported work?				
Yes	56.7%	70.7%	32.2%	73.5%
No	43.3%	29.3%	67.8%	26.5%

these work experience questions indicate a great deal of variation among the participants. Not surprisingly, nearly two thirds of the participants are female. The average age is 37.5 years. Slightly over half are Caucasian, and slightly less than one quarter are either African American or Hispanic. Only 7.6 percent have a disability. Nearly half (44 percent) of the participants have some college but not a college degree, nearly a quarter have an undergraduate degree, and 16 percent have either some graduate work or a graduate degree. Nearly 60 percent have neither a license nor a certificate. Finally, 43 percent report that they have not had prior training in supported work.

As discussed above, participants were randomly assigned groups of participants, rather than assigning individuals to one of the three experimental groups. Consequently, the characteristics of the members of the three experimental groups varied more than one would have expected had individuals, rather than groups, been assigned randomly. For example, had individuals been assigned to the experimental groups, it would have been much more likely to find the groups' racial composition and education to be closer in line with the findings presented in the "total" column, as is the case with all the other characteristics included in Table 1. Of the members of Group One, 50 percent are African

American, 72 percent of Group Two is Caucasian, and 40 percent of Group Three is Hispanic. Similarly, over half of the participants in Group Three have some college education, whereas 12 percent of Groups One and Two have graduate degrees. Finally, while nearly three quarters of Groups One and Two report having prior training in supported work, only one third of Group Two has prior training.

A Chi-square analysis was conducted to see if the groups differed significantly on selected characteristics in Table 1 that could influence their scores. The "number of years worked in rehabilitation" continuous data were recoded into five categories: 0-4, 5-9, 10-14, 15-19, and 20 or more. Similarly, the "number of years worked at your present job" continuous data were recoded into five categories: 0-1, 2-3, 4-5, 6-7, and 8 or more.

The evaluators find that there is no significant difference among the groups for number of years worked in rehabilitation ($\chi^2 = 9.54$, $df = 8$, $p = .299$), number of years worked at their current position ($\chi^2 = 9.525$, $df = 8$, $p = .300$), and education ($\chi^2 = 17.538$, $df = 10$, $p = .063$). As data in Table 1 suggest, the evaluators find a significant difference among the groups for race ($\chi^2 = 60.817$, $df = 10$, $p < .001$) and prior training in supported work ($\chi^2 = 36.125$, $df = 2$, $p < .001$).

Table 2. Findings from the t-test Analyses of the Six Hypotheses.

	$O_1 = O_2^*$		$O_3 > O_2$		$O_4 > O_2$		$O_4 > O_1$		$O_3 > O_1$		$O_3 > O_4$	
	O_1	O_2	O_3	O_2	O_4	O_2	O_4	O_1	O_3	O_1	O_3	O_4
n =	60	92	92	92	84	92	84	60	92	60	92	84
Mean	10.9	11.1	12.9	11.1	12.8	11.1	12.8	10.9	12.9	10.9	12.9	12.8
S.D	2.95	2.51	2.14	2.51	2.07	2.51	2.07	2.95	2.14	2.95	2.14	2.07
t statistic	-0.393*		5.439*		5.145*		4.426*		4.647*		0.282	
t critical, 1 tail	1.981*		1.654		1.654		1.660		1.660		1.654	
d.f.	112		177		172		99		99		173	

* Note: The t statistic and t critical values for the $O_1 = O_2$ hypothesis are for a 2 tail test; all others are for a 1 tail test.

* Indicates significance at the .001 level.

Knowledge Gain

A series of t-tests was conducted to analyze the hypotheses listed above, the results of which are contained in Table 2.

The first hypothesis ($O1 = O2$) tests the degree to which two of the three groups had an equal understanding of supported work prior to training. Participants' scores on the pretest are 10.9 (s.d. = 2.95) for Group One and 11.1 (s.d. = 2.51) for Group Two. Despite demographic differences indicated in Table 1 that suggest the groups could be different in important ways, these scores are not significantly different. The t-test for this hypothesis was conducted using a 2-tailed test, unlike the other hypotheses, so differences in either direction could be analyzed. Within Groups One and Two, participants' knowledge about supported work was not significantly different prior to attending the continuing education workshop.

Hypotheses two through five ($O3 > O2$; $O4 > O2$; $O4 > O1$; and $O3 > O1$) are intended to measure if participants learned the material presented during the workshop. As figures in Table 2 indicate, all of the hypotheses are supported at the .001 level. Participants' knowledge of supported work significantly improved after the workshop.

The final hypothesis ($O3 > O4$), that the Pretest-Posttest group's posttest scores ($O3$) would be significantly higher than the Posttest Only group's scores ($O4$), is the most important factor for this experiment, because it addresses the issue of testing. As the findings reported in Table 2 indicate, this hypothesis is not supported with the data. Group Two's posttest mean scores were 12.9 (s.d. = 2.14), and Group Three's posttest mean scores were virtually the same, 12.8 (s.d. = 2.07). Participants in Group Two who took a pretest did not score significantly higher on their posttest than those in Group Three who took the posttest only.

Conclusion

To maintain their professional licenses or certifications, social workers throughout the nation must

participate in continuing education. A necessary first step toward having continuing education improve workers' practice effectiveness is to increase participants' knowledge of the subject matter included in the sessions. A simple way to evaluate the impact continuing education has on knowledge gain is to use the one group pretest-posttest design.

Methodologists have long argued that this design has weak internal validity because evaluators are unable to rule out rival variables, such as "testing," in addition to the program, which may have influenced participants' scores on knowledge tests. This article reports the findings of an experiment that directly addresses the "testing" issue.

In this study, participants in twelve continuing education sessions were randomly assigned into one of three groups: Pretest Only, Pretest-Posttest, and Posttest Only. A series of t-tests was then conducted to compare participants' scores on an identical instrument designed to measure their knowledge of the subject matter. All of the results of the analyses are significant and concur with the expected direction that continuing education increases knowledge, with the exception of the hypothesis assessing the impact "testing" has on participants. If testing does indeed cause group scores to increase as theory suggests, the evaluators would have observed higher posttest scores for the Pretest-Posttest group, than posttest scores in the Posttest Only group. The data do not support this hypothesis. On the contrary, these scores are not different statistically and are virtually the same, 12.9 versus 12.8, suggesting no testing effect on the groups. It is possible, however, that testing does have an effect on individuals that is not detected in the group scores.

The ability to generalize these findings seems relatively strong. Selection of participants in the workshops was based primarily on agencies requiring their employees to attend. Very few voluntarily chose to attend, hence there is little danger of highly motivated individuals volunteering to attend

Using a Pretest-Posttest Design to Evaluate Continuing Education Programs

these workshops. Since the evaluators find testing to have no significant impact on groups' scores, there can be little or no interaction between testing and the workshop itself (see Campbell & Stanley, 1963). However, presenters should consider administering a pretest prior to the start of their presentation to learn both participants' knowledge of the topic and to direct group discussions.

These findings, while far from definitive, are important for continuing social work education. The findings suggest it is possible to use the simple pretest-posttest design to conduct meaningful evaluations of continuing education programs intending to increase participants' knowledge. Methodologists' concerns about the weaknesses of this design may be somewhat overemphasized.

Appendix

Questions used in the pretest and posttest (multiple choice answers omitted)

1. What is the overall purpose of supported employment?
2. There are two people working at the same job, for the same length of time, doing the same thing, one is disabled and the other is not. The person who is not disabled is paid \$7.00 per hour. How much should the person with the disability be paid?
3. Characteristics of an "Employment Specialist" [include which of the following?]
4. Which of the following are responsibilities of an "Employment Specialist"?
5. What is the main outcome for an employment specialist to seek for his or her clients?
6. Which of the following would be least helpful in developing a client profile?
7. Which one of the following is NOT one of the basic ideas of Supported Employment?
8. Once a client is hired, which of the following should ideally occur?
9. Which of the following is the best rationale for why we examine a consumer's domestic domain?
10. Which of the following is important to look for during a work site analysis?
11. Which of the following describes desirable characteristics of job coach intervention?
12. Which of the following is a good tip for working with families?
13. What should be your overall goal as an employment specialist?
14. [Which of the following is] An example of a natural support?
15. Which of the following statements [concerning supported work] is true?

Using a Pretest-Posttest Design to Evaluate Continuing Education Programs

References

- American Association of State Social Work Boards. (1996a). *Continuing education summary: 1996 edition*. Culpeper, VA: American Association of State Social Work Boards.
- American Association of State Social Work Boards. (1996b). *Social work laws and board regulations: A state comparison study*. Culpeper, VA: American Association of State Social Work Boards.
- Bibus, A. A. III & Rooney, R. H. (1995). Training child welfare practitioners in work with involuntary clients. *Journal of Continuing Social Work Education, 6(4)*, 13-22.
- Campbell, D. T. & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin Company.
- Cook, T. D. & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Boston: Houghton Mifflin Company.
- Denning, J. W. (1993). Statewide family preservation training: The staff response. *Journal of Continuing Social Work Education, 5(4)*, 5-10.
- Dietz, T. J. (1998). Incorporating follow-up to evaluate the impact continuing professional education programs have on social work practice. *Professional Development: The International Journal of Continuing Social Work Education, 1(1)*, 16-21.
- Feucht, T. E., Stephens, R. C., & Gibbs, B. H. (1991). Knowledge about AIDS among intravenous drug users: An evaluation of an education program. *AIDS Education and Prevention, 3(1)*, 10-20.
- Lindsey, E. W., Yarbrough, D. B., & Morton, T. D. (1987). Evaluating interpersonal skills training for public welfare staff. *Social Service Review, 61(4)*, 623-635.
- McCowan, R. J., McGregor, E. N., & LoTempio, J. (1982). Competency-based evaluation of social services training. *Journal of Continuing Social Work Education, 2(1)*, 11-13.
- Mueller, B. J. (1985). Participant action planning: A qualitative method to evaluate retention of learning. *Journal of Continuing Social Work Education, 3(2)*, 9-14.
- Neuman, W. L. (1991). *Social research methods*. Boston: Allyn and Bacon.
- Rooney, R. H. (1988). Measuring task-centered training effects on practice: Results of an audiotape study in a public agency. *Journal of Continuing Social Work Education, 4(4)*, 2-7.
- Royse, D. (1995). *Research methods in social work (2nd ed.)*. Chicago: Nelson-Hall Publishers.
- Rubin, A. & Babbie, E. (1997). *Research methods for social work (3rd ed.)*. Pacific Grove, CA: Brooks / Cole Publishing Co.