

4-1-2006

Real World Evaluation

Drew Lenore Betz

Washington State University, betz@wsu.edu

Laura Griner Hill

Washington State University, LauraHill@wsu.edu



This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 4.0 License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

Recommended Citation

Betz, D. L., & Hill, L. G. (2006). Real World Evaluation. *The Journal of Extension*, 44(2), Article 20.
<https://tigerprints.clemson.edu/joe/vol44/iss2/20>

This Research in Brief is brought to you for free and open access by the Conferences at TigerPrints. It has been accepted for inclusion in The Journal of Extension by an authorized editor of TigerPrints. For more information, please contact kokeefe@clemson.edu.



April 2006 // Volume 44 // Number 2 // Research in Brief // 2RIB9



Real World Evaluation

Abstract

We address the challenges of creating evaluation protocols that serve interests of both researchers and field faculty. WSU Extension used true and retrospective forms of pretest in evaluations of 100 adults attending the Strengthening Families Program for Parents and Youth 10-14 Years. We hypothesized that both forms of pretest would show positive change to posttest and that "Desirable" item types would show greater change. Both forms of the test indicated significant change in intervention-related behaviors, with greater change on retrospective tests and socially desirable items. We recommend using both a true pre-test and retrospective pre-test to satisfy researchers and practitioners.

Drew Lenore Betz

Extension Educator
Washington State University Whatcom County Extension
Bellingham, Washington
betz@wsu.edu

Laura Griner Hill

Assistant Professor and Extension Specialist
Washington State University Extension
Human Development Department
Pullman, Washington
laurahill@wsu.edu

Washington State University

Introduction

Washington State University Extension selected the Strengthening Families Program for Parents and Youth 10-14 Years (SFP), cited in the literature as the Iowa Strengthening Families Program (Molgaard, Kumpfer, & Fleming, nd), as a model program. We chose SFP for several reasons: 1) it addressed a population that was underserved (parents and preadolescents) in the state; 2) it was an Extension-developed best practice model that we were drawn to; and 3) we knew we could access funding for this particular program, because local funders had begun to require use of best-practice models.

Our goal was to encourage implementation of SFP as widely as possible across the state. We formed an internal Extension partnership among field-based Extension educators, the state Family Living Program Leader, and a member of the Human Development (HD) faculty whose focus is prevention research. The addition of the HD faculty member was a key element in planning the statewide initiative. The HD faculty member was familiar with the extensive research on SFP and was eager to partner with Extension to see it used more widely in Washington State.

An early decision was made by the team to conduct a solid evaluation to determine what adults and youth participants learned from the program as it was implemented in local areas without the benefit of a large research grant to fund the sites. A separate study was launched to look at patterns of implementation across the state.

The natural tension that exists between academic research practices used by human-science faculty in land-grant institutions and field-based evaluation conducted by youth and family

Extension professional was evident early in our process (Myers-Walls, 2000). We realized that the human science faculty want research that is rigorous and can withstand scrutiny by peer review. The goals of the human science researcher are linked to discovery, confirmation, replication, and dissemination. The youth and family Extension professionals want programs and practices that can be of maximum benefit to the participants, are effective, and can be implemented by a range of staff. The goals of the Extension professionals are engagement, transformation through education, effectiveness, and accountability.

Common ground between human science researchers and Extension professionals can be found in mutual concern about whether a program or practice is effective and is congruent with its design (i.e., delivering what it promises in the way it is designed). The common goals of assuring that programs are implemented with fidelity and are effective in the real world are especially important as pressure increases to adopt best practice models.

The development of the evaluation tools and protocol reflected some of the tensions described. Local SFP facilitators, who were partnering with WSU Extension but not part of the Extension system, were reluctant to ask the families to do too much paperwork at the start of the program. They were concerned that it would scare families away.

Extension educators were more comfortable with the need to do evaluation but were concerned about response-shift bias that has been reported when a true pre- and posttest design is used (Howard & Daily, 1979). Retrospective pre- and posttest evaluations had become more frequent in practice within the WSU system. Extension educators felt that their concerns were addressed by using the retrospective pretest format and appreciated the fact that this format engaged parents in reflecting about their own learning.

The prevention researcher was convinced that a traditional pre- and posttest design provides more valid results, because numerous cognitive biases, including social desirability bias, have been reported to influence retrospective ratings (Schwarz & Sudman, 1994). Both camps agreed that a solid evaluation was important and that having reliable results to report would benefit their programs. After much discussion and several pilot evaluations, we decided to examine social desirability bias by comparing traditional and retrospective pretest results. We also examined whether response-shift bias, resulting in negligible or negative change, was a problem with use of traditional pretest to posttest scores.

A major question of interest in setting up the evaluation was how the traditional pretest would compare to the retrospective pretest, both overall and by item type. In the original measure, all items were worded so that parents were asked to report how closely their parenting matched desirable parenting behaviors (e.g., "I follow through with consequences each time my youth breaks a rule"); this phrasing implicitly suggests that following through with consequences is the desired and normative behavior (Schwarz & Sudman, 1994). Therefore, we changed the wording on half the items so that raters were endorsing an undesirable parenting behavior or attitude rather than a desirable one (e.g., "It is hard to apply consequences consistently when my youth breaks the rules").

We expected that participants would have an easier time admitting to "Undesirable" behaviors if they were phrased this way. Because the "Desirable" items pull more strongly for socially desirable responses, we expected that parents would inflate the differences between retrospective pretest and posttest scores. If this were the case, we would see greater differences between traditional and retrospective scores on the Desirable items than on the Undesirable items.

Our hypotheses were as follows.

- Posttest scores would show improvement from both traditional and retrospective pretests
- Improvement from both forms of pretest to posttest would appear greater on items with socially desirable content (Desirable items)
- Differences between traditional and retrospective pretest ratings would also be greater on Desirable items.

Method

Measures

We adapted a measure of intervention-related parenting attitudes and behaviors included with the ISFP manual (Molgaard, Kumpfer, & Fleming, nd) by rephrasing half the items so that undesirable behaviors were presented as normative. Items were rated on a 7-point Likert-type scale, with anchors ranging from "Never" (1) to "Always" (7), with the midpoint (4) labeled "Half the time." We reverse-scored the Undesirable items for ease of comparison with the Desirable items.

Desirable Behavior Items

- I enjoy spending time with my youth.
- I give compliments and special rewards when my youth follows the rules.
- I feel like I know what my youth's dreams and goals are.

- I can wait to deal with problems with my child until I have cooled down.
- I let my youth know in advance what the consequences are for breaking rules.
- I am able to spend special time one-on-one with my youth
- My youth talks to me when he/she is upset.
- I give points and rewards when my child learns to follow a rule or do chores at home.

Undesirable Behavior Items

- When I am upset with my youth, I tend to blame and criticize him/her.
- It is hard to understand my youth's point of view.
- I lose my temper when my youth talks back.
- It is hard for me to show love to my youth.
- It is hard to enjoy being together and doing things as a family.
- Getting my youth to help with chores is a problem.
- Getting my youth to do homework is a problem.
- It is hard to apply consequences consistently when my youth breaks rules.

Sample

We collected data from 177 participants who attended 15 Iowa Strengthening Families Program for Parents and Youth 10 -14 Years (ISFP) series conducted over a 2-year period (January 2002 through December 2003). Out of the total sample, 100 participants completed evaluations both at beginning and end of program and completed all three measurements (traditional pretest, retrospective pretest, and posttest). Average pretest scores of those who completed all measurements did not differ significantly from average pretest scores of participants who did not complete posttests (i.e. who dropped out of the program).

Procedure

We used a within-subjects design for the present study, as is customary in studies of retrospective ratings (Henry, Moffitt, Caspi, Langley, & Silva, 1994; Ross, 1989; Wilson & Ross, 2001).

Traditional pretests were administered to participants the first night of the program. As is customary in many Extension evaluations, both retrospective pretests and posttests were administered on the final night of the program. Posttest and retrospective pretest items were on the same page. Providers instructed participants to cover the posttest rating column with a paper and to rate their behaviors and attitudes "THEN" (before the program), and then to cover the retrospective pretest ratings while rating their behaviors "NOW" (at end of program) on the posttest items.

Analysis

We calculated change scores by subtracting one score from another. For example, the difference between traditional pretest and posttest scores was calculated by subtracting the traditional from the posttest score. We used paired t-tests to test the significance of differences between traditional and retrospective pretest scores, and between both types of pretest and posttest scores.

Results

In Table 1, we present mean scores overall and by item type for all three measurements as well as change scores and their significance level.

Table 1.
Mean Scores and Change Scores Across Three Test Types, Overall and by Item Type

Item type	Pretest	Retrospective	Difference Between Pretest and Retrospective	Posttest	Change Pretest to Post	Change Retro to Post
All items	4.55 (0.66)	4.24 (0.80)	-0.31***	5.20 (0.58)	0.65***	0.96***
Desirable	4.58 (0.77)	4.10 (0.95) ^a	-0.49***	5.21 (0.73)	0.63***	1.11***
Undesirable	4.52 (0.74)	4.39 (0.94) ^a	-0.13* *	5.19 (0.71)	0.68***	0.81***

 $p < .001$

^a indicates that Desirable and Undesirable column values are significantly different from each other ($p < .001$)

Hypothesis 1: Improvement from Pretest to Posttest

As indicated in Table 1, mean scores across all items showed significant change from both traditional and retrospective pretests to posttest. Significant change was observed regardless of item type (i.e., Desirable and Undesirable) or pretest form (traditional pretest and retrospective pretest).

Hypothesis 2: Differences by Item Type

Average change was greater from pretest to posttest for Desirable than for Undesirable items. The Desirable-Undesirable difference was larger for retrospective ratings than for traditional pretest ratings.

Hypothesis 3: Differences Between Traditional and Retrospective Pretest Ratings

Retrospective pretest scores were significantly higher than true pretest scores for Desirable items. However, pretest and retrospective ratings for Undesirable items were equivalent.

Discussion

All three of our hypotheses were validated in this study. Notable findings included the following.

- Parents reported improvement over time, whether measured by true pretest or retrospective pretest. These results reflect the integrity of both the program and the evaluation and are satisfying to field staff, community partners, and researchers.
- Social desirability is influential when using either form of pretest, but the differences are magnified when using the retrospective version of the evaluation, especially on the "Desirable" items (i.e., items that pull for socially desirable answers).

When allowed to rate both "then" and "now" aspects of their behavior, adults in our study reported greater changes when the items asked about behaviors that are viewed as positive or desirable. On the other hand, and consistent with the idea that social desirability bias plays a role in self-ratings, ratings of the "Undesirable" items were less inflated, in fact showing no significant differences between the two forms of pretest.

Response shift bias may also have played a part in the differences just discussed. Adults were less sure of their knowledge base prior to taking the course and re-evaluated their own learning differences on retrospective form of the evaluation. Although the differences in the Undesirable pretest and retrospective pretest items were not significant, the direction of the difference is what might be expected from a response-shift bias. The mean score was slightly lower for the retrospective test on these items.

Implications for Practice

There are several implications of this study's findings.

Goals Can Be Balanced

First, the study shows that goals of the research and field-based evaluation can be balanced in real-life programs.

All parties in our study were satisfied with the outcome. The evidence that both types of pretests showed significant differences was helpful to Extension educators and SFP facilitators and contradicts the common wisdom that traditional pretests will not show change. The evidence of higher scores on the Desirable items on the retrospective pretest leads us to speculate that it is important to consider effects of social desirability bias as well as response-shift bias when designing survey items, especially if they are to be used in retrospective pretests.

Both Forms Are Useful

Second, both forms of the pre-test are useful, and we would recommend use of both to colleagues in the Extension field who are partnering with researchers in their respective institutions.

The traditional pretest-posttest model gives the validity we seek with funders and the assurance we need to demonstrate our outcomes with our colleagues who are used to more rigorous forms of research used by plant and animal sciences as well as social sciences.

On the other hand, whereas traditional pretests may provide a more accurate picture of mean levels of group change (i.e., program effects), retrospective pretests provide a more accurate picture of how much participants feel they have benefited from the program individually.

If goals of the evaluation are solely to demonstrate program effectiveness, we recommend use of a traditional pretest-posttest design. If goals of the evaluation are solely to assist parents' reflection of their experience or to assess their perceptions of change, we recommend use of the retrospective pretest-posttest design.

It's Important to Engage All Partners

Third, it is important to engage all partners in the evaluation process.

In our case we had community partners, non-Extension community staff, research faculty, Extension professionals, and adult participants to satisfy. Personal experience has taught us that the evaluation process can break down with any of these parties. A well-thought-out evaluation is only one element of the process. Proper training in evaluation protocol, agreement to give time to the evaluation during the program by the program staff, and follow-through in delivering the evaluations to the researcher are also keys to the success of the effort.

We make a special effort to bring our researcher to each of our WSU-sponsored facilitator trainings. Potential facilitators are educated on the hows and whys of our study and are encouraged to participate. Presentation of the rationale behind and advantages of evaluation helps facilitators to perceive the evaluation as a positive aspect of program delivery. For example, trainers suggest that facilitators use the retrospective pretest as a tool to help parents appreciate the skills they have learned in the program. The success of our approach is evident in the numerous sites that choose to submit evaluation data.

Completing the feedback loop by returning reports to program staff and community partners is another key element. Community partners find evaluation reports useful in applying for continued funding, Extension educators find them useful in completing annual activity reports, and campus faculty find them helpful in examining program effectiveness and in publishing research reports. All parties find evaluation data useful in monitoring and improving program delivery.


Conclusion

In conclusion, it is possible to create real-world evaluation systems that satisfy the needs of researchers, field staff, and community partners. The results of the study reported here have been beneficial for our efforts to both create consistency in statewide evaluation of our program and in finding formats that bring good information to all involved. Extension can provide leadership for sound evaluation practices that can be used beyond the circle of its own programs.

References

- Henry, B., Moffitt, T. E., Caspi, A., Langley, J., & Silva, P.A. (1994). On the "remembrance of things past": A longitudinal evaluation of the retrospective method. *Psychological Assessment, 6*(2), 92-101.
- Howard, G. S., & Dailey, P. R. (1979). Response-shift bias: A source of contamination of self-report measures. *Journal of Applied Psychology, 64*(2), 144-150.
- Molgaard, V. K., Kumpfer, K., & Fleming, E. (nd). *Strengthening families program for parents and youth 10-14: A video based curriculum (Leader guide)*. Ames, Iowa: Iowa State University Extension.
- Myers-Walls, J. A. (2000). An odd couple with promise: Researchers and practitioners in evaluation settings. *Family Relations: Interdisciplinary Journal of Applied Family Studies, 49*, 341-347.
- Redmond, C., Spoth, R., Shin, C., & Lepper, H. S. (1999). Modeling long-term parent outcomes of two universal family-focused preventive interventions: One-year follow-up results. *Journal of Consulting & Clinical Psychology, 67*(6), 975-984.
- Ross, M. (1989). Relation of implicit theories to the construction of personal histories. *Psychological Review, 96*(2), 341-357.
- Schwarz, N., & Sudman, S. (1994). *Autobiographical memory and the validity of retrospective pretests*. New York: Springer Verlag.
- Spoth, R., Gyll, M., Chao, W., & Molgaard, V. (2003). Exploratory study of a preventive intervention with general population African American families. *Journal of Early Adolescence, 23*,

Wilson, A. E., & Ross, M. (2001). From chump to champ: People's appraisals of their earlier and present selves. *Journal of Personality & Social Psychology*, 80(4), 572-584.

Copyright  by Extension Journal, Inc. ISSN 1077-5315. Articles appearing in the Journal become the property of the Journal. Single copies of articles may be reproduced in electronic or print form for use in educational or training activities. Inclusion of articles in other publications, electronic sources, or systematic large-scale distribution may be done only with prior electronic or written permission of the Journal Editorial Office, joe-ed@joe.org.

If you have difficulties viewing or printing this page, please contact JOE Technical Support