

Testing the Retrospective Pretest with High School Youth in Out-of-School Time Programs

Jill Young

After School Matters
jyy001@gmail.com

Leanne Kallemeyn

Loyola University Chicago
lkallemeyn@luc.edu


Abstract

Practitioners and evaluators face several constraints in conducting rigorous evaluations to determine program effect. Researchers have offered the retrospective pretest/posttest design as a remedy to curb response-shift bias and better estimate program effects. This article presents an example of how After School Matters (ASM) tested the use of retrospective pretest/posttest design for evaluating out-of-school time (OST) programs for high school youth participants. Differences between traditional pretest and retrospective pretest scores were statistically significant, but effect sizes were negligible, indicating that both pretests yielded similar results. Interviews with youth led to 3 key findings that have implications for ASM using retrospective pretests with youth: response-shift bias was more prominent in youth interviews than in quantitative findings, youth recommended reordering the questions so that the retrospective pretest appears first to increase comprehension, and acquiescence bias emerged in the interviews. This study demonstrates that the retrospective pretest/posttest design can be an alternative to the traditional pretest/posttest design for OST at ASM. These findings are important for ASM and other youth-serving organizations, which often have limited capacity to survey youth multiple times within 1 program session.

Key words: out-of-school time, high school, program evaluation, survey, pretest, response-shift bias

Introduction

Practitioners and evaluators face several constraints in conducting rigorous evaluations to determine program effect (Bamberger, Rugh, Church, & Fort, 2004). These constraints include limited time, expertise, leadership support, and budget (Reed & Morariu, 2010); inaccessible or

 New articles in this journal are licensed under a Creative Commons Attribution 4.0 License. This journal is published by the [University Library System](#), [University of Pittsburgh](#) and is cosponsored by the [University of Pittsburgh Press](#). The Journal of Youth Development is the official peer-reviewed publication of the [National Association of Extension 4-H Agents](#) and the [National AfterSchool Association](#).

Retrospective Pretest in OST Programs

incomplete data (Allen & Nimon, 2007); and competing priorities between funders and program providers (Benzies, Clarke, Barker, & Mychasiuk, 2012). Out-of-school time (OST) programs experience similar demands and constraints. These practical problems often result in evaluators using designs that require minimal resources, such as traditional pretest/posttest designs and retrospective pretest/posttest designs. This article presents an example of how one organization tested the use of retrospective pretest/posttest design for evaluating OST programs for high school youth participants.

In a retrospective pretest/posttest design, program participants are asked to rate themselves on a variable of interest based on how they feel currently, and then they are asked to rate themselves on that same variable based on how they felt at the beginning of the program (Howard, Ralph, et al., 1979). The pretest and posttest responses are typically collected at the same time. Self-report data, which are typically used in this design, are common in youth research because such data capture youth voice (Durlak, Weissberg, & Pachan, 2010). Retrospective pretests are practical for several reasons, although they are an imperfect design (Lamb, 2005). This design circumvents constraints OST programs experience when evaluating programs. Retrospective pretests take less time to administer than traditional pretests and create less of a burden for respondents. They reduce attrition and missing data. They can be useful when traditional pretests are not possible for logistical or other reasons. In addition, they can reduce response-shift bias (Lamb, 2005), and avoid introducing confusing terms before participants are ready for them.

The retrospective pretest/posttest design has been tested and explored in several studies to measure perceived changes in behaviors or attitudes of respondents (Cantrell, 2003; Hill & Betz, 2005; Howard, Ralph et al., 1979; Moore & Tananis, 2009; Nimon, Zigarmi, & Allen, 2011; Pelfrey & Pelfrey, 2009). Furr and Bacharach (2014) stated that psychometric properties such as reliability and validity are sample-dependent: the characteristics of the survey respondents and the contexts in which they complete the survey matter. Several studies have tested the design with youth (Bobilya & Faircloth, 2017; Moore & Tananis, 2009; Sibthorp, Paisley, Gookin, & Ward, 2007).

This article presents the process After School Matters (ASM) completed to determine if the design was appropriate for its programs and evaluation needs. ASM is a nonprofit organization that provides OST programs to Chicago public high school youth. Programs focus on project-based learning and provide youth with skills for college, career, and beyond. There are 1,500 programs and 26,000 opportunities at over 400 different sites in Chicago. These programs are

offered during three program sessions each year, with each session serving between 7,000 and 13,000 youth. Fall and spring sessions run approximately 10 weeks. ASM represents the largest OST program provider for high school youth in one of the largest cities in the United States.

Cognitive Processes and Biases in Self-Report

Howard (1980) acknowledged that all self-report instruments, including those used in traditional pretests, retrospective pretests, and posttests, are prone to biases and thus threaten internal validity. The fallibility of self-reported measures is primarily related to the complex cognitive process respondents engage in to answer the questions. Whether and to what degree respondents complete this process is pivotal to the validity of the information collected through self-report measures. According to Schwarz (1999), the process includes five steps: understanding the question, recalling relevant behavior, making inferences and estimations, selecting a response, and editing answers. Optimizing occurs when respondents successfully complete this process (Krosnick, 1999). Often, respondents are not motivated to engage in the full cognitive process throughout the survey. In these situations, respondents adapt their response strategy in what Krosnick called satisficing. Satisficing is more likely to occur the greater the task difficulty, the lower the respondent's ability, and the lower the respondent's motivation.

Cognitive functioning varies depending on the survey respondent's age (Borgers, Sikkel, & Hox, 2004). According to de Leeuw (2011), cognitive functioning is well developed by the time youth reach adolescence at age 12. Youth of this age follow the same cognitive steps as adults in responding to survey questions, but researchers must pay additional attention to certain steps. While adolescents' memory capacity is fully developed, their memory speed is not, so youth may require more time to respond to questions that require recall. De Leeuw noted that adolescents need about 1.5 times as much time as an adult to process information. Finally, de Leeuw reported that youth ages 12 and older are sensitive to peer pressure and group norms, and advised researchers to survey youth confidentially and remind them that there are no correct answers.

Several biases pose a threat to construct validity when using self-report measures, including the retrospective pretest/posttest design (Hill & Betz, 2005; Howard, 1980; Nimon et al., 2011; Ross, 1989; Schwarz & Oyserman, 2001; Taylor, Russ-Eft, & Taylor, 2009). First, response-shift bias occurs when survey respondents overestimate or underestimate themselves at pretest because they do not have an adequate understanding of the construct on which they are

Retrospective Pretest in OST Programs

evaluating themselves—the knowledge, skills, and attitudes that the program intends to affect (Lam & Bengo, 2003). This often leads to inaccurate estimates of program impact, especially in training or education programs when the intent is to change a participant's understanding or awareness of a particular construct (Allen & Nimon, 2007; Howard, 1980).

Second, social desirability is a bias in which respondents over-report more socially accepted attitudes and behaviors, and under-report those that are less socially accepted (Krosnick, 1999). Acquiescence bias is "the tendency to endorse any assertion made in a question, regardless of its content," also referred to as "yea-saying or nay-saying" (Krosnick, 1999, p. 552). It is more common among people with limited cognitive skills or less cognitive energy. It is also more common when the question is difficult or ambiguous, respondents are encouraged to guess, or after respondents have become fatigued. Effort justification bias describes when a participant who did not find the intervention particularly effective alters his or her responses in retrospective assessment to exaggerate change and justify the investment he or she has made (Nimon, Zigarmi, & Allen, 2011). In implicit theory of change, participants assume the intervention had its desired effect (Ross, 1989). Finally, recall bias is the distortion or degradation of memory (Hill & Betz, 2005; Ross, 1989). These biases affect the validity of the retrospective pretest/posttest design.

Detecting biases can be difficult. Furr and Bacharach (2014) indicated that though some indexes are available to detect social desirability or acquiescence, it is preferred to discourage self-report biases on the front end. The literature offers several possible solutions for doing so. Krosnick (1999) advised encouraging respondents to think carefully before answering questions, and making instruments anonymous or confidential to encourage honest responses. Schwarz and Oyserman (2001) suggested evaluators answer the survey questions themselves first and pilot the survey before implementing it widely. Moore and Tananis (2009) recommended incorporating open-ended questions and focus groups to provide clarity around responses and bolster evidence of validity. In addition to these suggestions, Furr & Bacharach advised evaluators to minimize respondent fatigue, distraction, or frustration; write simple, neutral items; provide forced or limited choices to avoid extreme responses; and write balanced scales. This study explores the retrospective pretest in relation to potential biases for youth participating in ASM.

Methods

Participants

All ASM youth that registered for the fall 2015 program session were invited to complete the traditional pretest. There were 6,574 youth who completed the program in fall 2015. The youth included in this population were all high school students, ages 13 to 19 and grades 9 through 12. The gender breakdown was similar to previous sessions, with 61% female, 39% male, and the remaining youth chose not to identify. The racial and ethnic breakdown was 56% Black/African-American, 33% Hispanic/Latinx, 5% two or more races, 3% Asian, and 3% Caucasian. Youth in ASM programs are typically from lower socioeconomic backgrounds, with 87% of youth who receive free or reduced lunch in their schools.

The study included only youth who completed programs and completed all survey items for the traditional pretest, posttest, and retrospective pretest questions, resulting in a sample size of 4,311 (65.6% response rate). There were no major demographic differences between the youth who completed all items on all three tests, the youth who did not complete all surveys and items, and the overall population of youth enrolled.

Additionally, 30 youth also participated in interviews. Youth were purposively selected at the program level using cluster sampling to ensure representation of ASM regions, content areas, youth demographic characteristics, and previous participation patterns, including program attendance and completion rates.

Instrumentation

ASM was most interested in demonstrating changes in specific 21st century skills; more specifically, leadership, teamwork, problem solving, public speaking and oral communication, meeting deadlines, and accepting constructive criticism. These skills were chosen based on literature on 21st century skills for youth (Forum for Youth Investment, 2010), commonly reported skills in ASM instructors' weekly plans, and teens' self-report open-ended comments about skills they learned in programs. These items were not exhaustive of all relevant 21st century skills. Youth were asked to rate themselves on each of the skills using a scale of 1 to 5. The survey included 17 questions total, many of which included several items. The traditional pretest 21st century skills items had a Cronbach's alpha of 0.87, the posttest Cronbach's alpha was 0.91, and the retrospective pretest Cronbach's alpha was 0.89, indicating high internal reliability of the items as a scale.

In the interviews, youth were asked to think aloud as they answered survey questions. Youth were also asked to share information about their program experience and the skills they gained because of the program. Additionally, interviews provided the opportunity for youth to reflect and provide feedback on completing the traditional and retrospective pretest questions, including the accuracy of their responses, their understanding of concepts or terms described in the survey items, and why their scores might have or might not have changed between the two survey administrations.

Procedures

Youth were given a traditional pretest as part of their application for fall 2015. All surveys were administered online through SurveyMonkey for the retrospective pretest and posttest during the last two weeks of the program as part of a post-program survey. Posttest questions appeared first in the survey, and retrospective pretest questions appeared second on a separate page. This procedure followed Schwarz's (1999) recommendation to reduce biases related to implicit theory of change, effort justification, and social desirability. The survey directions stated that answering the questions was optional, that youth responses would not affect their current or future participation in ASM programs, and that their responses were confidential and would not be shared outside of the research and evaluation team unless there was a safety concern. The directions also requested that youth answer the questions honestly.

Interviews took place during the last two weeks of programs. Each interviewer began the conversation by explaining that the intention of the interview was to collect honest feedback about the survey to make it easier for youth to complete. Interviewers explained to youth that their responses were confidential, and their responses would not be connected to their name. As part of the consent process, youth were also told that they could stop the interview at any time.

Data Analysis

Data analysis included descriptive statistics of the sample and reliability estimates on the scores from the traditional pretest, retrospective pretest, and posttest; a two-tailed dependent sample t-test to determine differences between traditional pretest and retrospective pretest scales; and effect sizes to determine the magnitude of the differences.

This study followed Merriam's (2009) process for qualitative data analysis. Qualitative data analysis incorporated both deductive and inductive approaches. Data from interviews were categorized by themes related to the five steps of the cognitive process and evidence of self-report biases, but the analysis also allowed for other categories to emerge. To ensure reliability and validity, this study incorporated triangulation of investigators, data sources, and methods (Merriam, 2009). The evaluators that assisted in data collection also held a peer review meeting and reviewed the audit trail.

Results

The traditional pretest and retrospective pretest means at the scale level were close at 4.06 and 4.01, respectively. The mean for the posttest was 4.21. A two-tailed dependent samples t-test was used to compare the average scale ratings in the traditional pretest and the retrospective pretest. There was a significant difference between the traditional pretest ($M = 4.06$, $SD = 0.73$) and the retrospective pretest ($M = 4.01$, $SD = 0.76$); $t = 3.031(4310)$, $p = .002$. The effect size was calculated using Cohen's d to determine the mean difference between the ratings at traditional pretest and retrospective pretest. The effect size was 0.06. Additionally, demographic differences were examined but none were found.

Interviews with youth led to three key findings that have implications for ASM using retrospective pretests with youth: response-shift bias was more prominent in youth interviews than in quantitative findings; youth recommended reordering the questions so that the retrospective pretest appears first to increase comprehension; and acquiescence bias emerged in the interviews.

First, response-shift bias was prominent in youth feedback collected through qualitative methods. Nearly two-thirds (63.3%) of the 30 youth interviewed provided evidence of response-shift bias. They commonly reported trouble with answering their traditional pretest questions. For some youth their traditional pretest scores were inaccurate because the way they evaluated the skills in question changed after completing their program. One youth noted:

I think this one [retrospective pretest] was more accurate because then you realize how much you improved at the time. 'Cause in the first one [traditional pretest] you already know like, 'oh, I think I'm so good at this, I'm so good at that,' but then after you see yourself improve, you're like, you think, 'I wasn't that good as how I am now.'

Retrospective Pretest in OST Programs

This youth felt she overrated herself in the traditional pretest because she did not understand the degree to which she could improve.

Youth were asked whether they understood the skills on which they rated themselves at the beginning and end of programming and whether their understanding of those skills changed. Several students reported such a change, especially for skills related to leading or working in teams. One youth shared that she thought she understood the skills in question, but the meaning of those skills changed after she completed her program:

When I first took it, I thought I did [understand], but when we say work well with teams in group projects . . . I thought about math and like huddling together in a circle discussing. . . . But when you get into After School Matters, it's not really that . . . you walk around and discuss different artists and techniques and give each other feedback.

This youth associated group projects with math problems at school, and she had difficulty translating what that might look like in her OST program. Another example was problem solving, which several youth reported they interpreted specifically as math problems in the traditional pretest. One youth shared, "The first thing that popped up was like math and English, stuff like that. What do you mean by solving problems? Life problems? Math problems? Problems with other people in the class?" In these examples, after youth completed the program, their definitions of the skills broadened as a result of their experience.

Second, youth generally understood the retrospective pretest question, but noted that it took them longer to complete because it followed the posttest question rather than preceding it. Researchers recommended that the posttest question appear before the retrospective pretest question to reduce the potential of effort justification bias and implicit theory of change (Howard, 1980; Sprangers & Hoogstraten, 1989; Taylor, Russ-Eft, & Taylor, 2009). However, nearly all youth (93.3%) interviewed reported the order of the posttest question and retrospective pretest question was confusing for them or could be confusing for other youth and recommended switching the questions so that the retrospective pretest question appeared before the posttest question.

Third, acquiescence emerged as a central bias. Four of the 30 youth interviewed provided evidence of acquiescence. Twenty youth complained in interviews that the survey was too long, and some youth admitted they did not take the time to respond thoughtfully. One youth shared,

Retrospective Pretest in OST Programs

“The survey gets long and boring so we’ll just go straight to this and answer it,” indicating that he did not read question stems after a certain point in the survey.

Based on this feedback, the decision was made to reexamine the quantitative data to investigate whether acquiescence was present and perhaps masking response-shift bias. Youth who selected the same response for all six items for the posttest question and all six items for the retrospective pretest were identified, resulting in the removal of 28.1% of respondents. An independent sample *t*-test was conducted to determine if the group that potentially acquiesced was different from the rest of the sample. There was a significant difference in average change between pretests between the “acquiescence” group ($M = -0.25$, $SD = 0.99$) and the “non-acquiescence” group ($M = 0.17$, $SD = .91$); $t = -12.91(2250)$, $p < .001$. The effect size for this analysis was 0.44, which is considered a small to moderate effect size (Howell, 2010).

Based on the finding that the two groups were indeed different, the two-way dependent sample *t*-test was re-run for the “non-acquiescence” group to detect the presence of response-shift bias. The results were more congruent with literature on response-shift bias. There was a significant difference between the traditional pretest ($M = 4.02$, $SD = 0.68$) and the retrospective pretest ($M = 3.85$, $SD = 0.71$); $t = 10.53(3097)$, $p < .001$. The effect size was 0.25, which is considered a small effect size (Howell, 2010).

The results of this analysis of the data should be interpreted with caution. There is no way to determine without additional data collection whether youth truly were acquiescing in their responses, or they simply did not feel they changed at all. Five youth who were interviewed also responded the same way to all 12 items. Three of these youth provided evidence of potential acquiescence, but the other two truly felt they were already strong in those skills and had not changed.

There was minimal evidence related to other self-report biases. One argument against the retrospective pretest is that its use may reduce response-shift bias but increase effort justification bias or implicit theory of change. If this were the case, one would expect decreases in the ratings from traditional pretest to retrospective pretest. However, though the change between the ratings at both time periods was significantly different, they were not substantially different (4.06 compared to 4.01, respectively). No evidence of effort justification emerged in the interview data either; however, two youth provided potential evidence of implicit theory of change. In both cases, youth suggested they must have changed because of their program but could not elaborate on the specific changes they experienced as it related to the skills.

Discussion

This study provides an example of how an organization might ensure a valid measure, given constraints and limited resources. The traditional pretest scores were significantly higher than the retrospective pretest scores as a scale, indicating youth overestimated their skill levels in the traditional pretest. Despite the statistical significance, the effect size was negligible, indicating that both pretest administrations yielded similar results. However, the findings from qualitative interviews with youth provided evidence that response-shift bias was present, with many youth reporting they did not understand what the skills in question meant within the context of their program until after they participated in it. Additionally, the qualitative data provided valuable insight into the administration procedures and other potential biases when using the retrospective pretest/posttest design with youth. Youth recommended switching the order of the pretest and posttest questions to increase the speed of comprehension for completing the survey. Also, acquiescence emerged as a potential bias, likely due to survey fatigue or confusion about the pretest question.

These findings are important for ASM and other youth-serving nonprofit organizations, which often have limited capacity to survey youth multiple times within one program session. After careful consideration of the study results and the practical problems ASM has experienced with evaluation, ASM decided to implement the retrospective pretest/posttest design, switch the order of the pretest and posttest questions, and shorten the survey to decrease acquiescence.

The retrospective pretest/posttest design may be most useful for programs that aim to evaluate self-reported perceived changes, have limited capacity or resources for evaluation, are concerned about overburdening youth and staff with data collection, and have evidence of response-shift bias. Other options for addressing response-shift bias include: defining constructs at traditional pretest (Howard, Schmeck, & Bray, 1979), indicating to survey participants that they will be assessed on the constructs (Hoogstraten, 1982), using an objective measure to assess the constructs (Howard, Schmeck, & Bray, 1979), and including a social desirability scale (Furr and Bacharach, 2014). But as Sibthorp et al. (2007) note, such alternatives are not always possible because they may be too difficult, burdensome, or costly.

This study also found acquiescence to be a potential bias when using the design with youth. This finding is consistent with that of Lam and Bengo (2003), who also suspected the presence of satisficing in their study. Although many of the studies hypothesized that the presence of other biases could have threatened internal validity with overinflated participant self-reports of

Retrospective Pretest in OST Programs

change, there is evidence that acquiescence masked the existence of response-shift bias in this study, likely due to survey fatigue or confusion about the retrospective pretest question. Survey respondents may satisfice and introduce other biases if they are fatigued by a long survey or they do not understand the question (Krosnick, 1999; Schwarz & Oyserman, 2001). Several youth in this study reported these issues.

Organizations should also consider the cognitive functioning of youth and which biases are most important to minimize. While the posttest question often precedes the retrospective pretest question (Howard, Schmeck, & Bray, 1979), switching the question order may increase comprehension for youth, who need more time to process, understand, and respond to surveys than adults (de Leeuw, 2011). Having the pretest question follow the posttest question may decrease implicit theory of change and effort justification bias, but it may increase acquiescence due to the extra cognitive effort it takes youth to understand. Therefore, both the use of the retrospective pretest question and the order in which it appears provide bias tradeoffs rather than perfect solutions.

Other biases may be a factor for youth at ASM, such as social desirability, implicit theory of change, effort justification, or recall (Hill & Betz, 2005; Howard, 1980; Nimon, Zigarmi, & Allen, 2011; Ross, 1989; Schwarz & Oyserman, 2001; Taylor, Russ-Eft, & Taylor, 2009). These biases present alternative explanations to the findings and may have affected results. However, open-ended responses from the survey and the information collected through interviews primarily supported evidence of response-shift bias and acquiescence. Future research on using the retrospective pretest/posttest design with youth should explicitly examine the design's effect on additional biases.

Several studies indicate the retrospective pretest scores are more highly correlated with related objective measures, which indicates concurrent validity (Bray, Maxwell, & Howard, 1984; Howard, Ralph et al., 1979; Hill & Betz, 2005). This study did not include an objective measure to determine whether traditional pretest or retrospective pretest measures are more highly correlated with objective measures of the same construct. It also did not incorporate cognitive interviews for the traditional pretest. Such evidence would further address which pretest administration is most accurate for youth.

This study demonstrates that the retrospective pretest/posttest design can be an alternative to the traditional pretest/posttest design for OST at ASM. The design is a practical and useful design for evaluators with limited capacity or resources to gather meaningful data that

Retrospective Pretest in OST Programs

accurately assess whether the program had its intended effects in a way that is quick, inexpensive, unobtrusive to staff and participants, and takes less time away from programs. Other programs in similar contexts may find the lessons learned using a retrospective pretest/posttest design at ASM relevant in their contexts. More importantly, this study outlines processes and considerations for using this design and self-report measure to evaluate OST programming.

References

- Allen, J. M., & Nimon, K. (2007). Retrospective pretest: A Practical technique for professional development evaluation. *Journal of Industrial Teacher Education, 44*(3), p. 27-42. Retrieved from <http://scholar.lib.vt.edu/ejournals/JITE/>
- Bamberger, M., Rugh, J., Church, M., & Fort, L. (2004). Shoestring evaluation: Designing impact evaluations under budget, time and data constraints. *American Journal of Evaluation, 25*(1), p. 5-37. doi:10.1177/109821400402500102
- Benzies, K., Clarke, D., Barker, L., & Mychasiuk, R. (2012). UpStart parent survey: A New psychometrically valid tool for the evaluation of prevention-focused parenting programs. *Maternity and Child Health Journal, 17*, p. 1452-1458. doi:10.1007/s10995-012-1152-2
- Bobilya, A.J., & Faircloth, W.B. (2017). Exploring course outcomes utilizing a new Outward Bound outcomes instrument. *Research in Outdoor Education, 15*, p. 114-136.
- Borgers, N., Sikkel, D., & Hox, J. (2004). Response effects in surveys on children and adolescents: The effect of number of response options, negative wording, and neutral mid-point. *Quality & Quantity, 38*, p. 17-33.
- Bray, J. H., Maxwell, S. E., & Howard, G. S. (1984). Methods of analysis with response shift bias. *Educational and Psychological Measurement, 44*, p. 781-804. doi:10.1177/00131644844444002
- Cantrell, P. (2003). Traditional vs. retrospective pretests for measuring science teaching efficacy beliefs in preservice teachers. *School Science and Mathematics, 103*(4), p. 177-185. Retrieved from <http://www.ssma.org/publications>
- de Leeuw, E. D. (2011). *Improving data quality when surveying children and adolescents: Cognitive and social development and its role in questionnaire construction and pretesting*. Report to the Annual Meeting of the Academy of Finland: Research Programs Public Health Challenges and Welfare of Children and Young People May 10-12 2011 in Finland.
- Durlak, J., Weissberg, R., & Pachan, M. (2010). A meta-analysis of after-school programs that seek to promote personal and social skills in children and adolescents. *American Journal of Community Psychology, 294*-309. Retrieved from: http://www.flume.com.br/pdf/Durlak_A_meta-analysisof_after_school.pdf

Retrospective Pretest in OST Programs

- Forum for Youth Investment. (2010). Readiness: The nation focuses on college and career preparations. *The School Administrator*, 6(67). Retrieved from <http://www.forumfyi.org/files/School%20Administrator%20Article%206-10.pdf>
- Furr, R. M., & Bacharach, V.R. (2014). *Psychometrics: An introduction*. Thousand Oaks, CA: Sage.
- Hill, L. G., & Betz, D. L. (2005). Revisiting the retrospective pretest. *American Journal of Evaluation* 26, p. 501-507. doi:10.1177/1098214005281356
- Hoogstraten, J. (1982). The retrospective pre-test in an educational training context. *Journal of Experimental Education*, 50(4), p. 200-204. Retrieved from <http://www.jstor.org/stable/20151460>
- Howard, G. S. (1980). Response shift-bias: A Problem in evaluating programs with pre/post self-reports. *Evaluation Review* 4, p. 93-106. doi:10.1177/0193841X8000400105
- Howard, G. S, Ralph, K. M., Gulanick, N. A., Maxwell, S. E., Nance, D. W., & Gerber, S. K. (1979). Internal validity in pretest-posttest self-report evaluations and a re-evaluation of retrospective pretests. *Applied Psychological Measurement* 3, p. 1-23. doi:10.1177/014662167900300101
- Howard, G. S., Schmeck, R. R., & Bray, J. H. (1979). Internal validity studies employing self-report instruments: A Suggested remedy. *Journal of Educational Measurement* 16, p. 129-135. Retrieved from <http://www.jstor.org/stable/1434456>
- Howell, D. C. (2010). *Statistical methods for psychology*. Belmont, CA: Wadsworth, Cengage Learning.
- Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, 50(1), p. 537-567. Retrieved from: <http://www.annualreviews.org/journal/psych>
- Lam, T. C. M., & Bengo, P. (2003). A Comparison of three retrospective self-reporting methods of measuring change in instructional practice. *American Journal of Evaluation* 24, p. 65-80. doi:10.1177/109821400302400106
- Lamb, T. (2005). The Retrospective pretest: An Imperfect but useful tool. *The Evaluation Exchange* 11(2), p. 18. Retrieved from <http://www.hfrp.org/var/hfrp/storage/original/application/d6517d4c8da2c9f1fb3dffe3e8b68ce4.pdf>
- Merriam, S. B. (2009). *Qualitative research*. (3rd ed.). San Francisco, CA: Jossey-Bass.
- Moore, D., & Tananis, C. A. (2009). Measuring change in a short-term educational program using a retrospective pretest design. *American Evaluation Association* 30(2), p. 189-202. doi:10.1177/1098214009334506
- Nimon, K., Zigarmi, D., & Allen, J. (2011). Measures of program effectiveness based on retrospective pretest data: Are all created equal? *American Journal of Evaluation* 32(1), p. 8-28. doi:10.1177/1098214010378354
- Pelfrey, W. V., Sr. & Pelfrey, W. V., Jr. (2009). Curriculum evaluation and revision in a nascent field: The utility of the retrospective pretest-posttest model in a homeland security program of study. *Evaluation Review* 33, p. 54-82. doi:10.1177/0193841X08327578

Retrospective Pretest in OST Programs

- Reed, E., & Morariu, J. (2010). *State of evaluation: Evaluation practice and capacity in the nonprofit sector*. Retrieved from http://www.innonet.org/client_docs/innonet-state-of-evaluation-2010.pdf
- Ross, M. (1989). Relation of implicit theories to the construction of personal histories. *Psychological Review*, *96*(2), p. 341-357. Retrieved from <http://www.apa.org/pubs/journals/rev/>
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, *54*(2), p. 93-105. Retrieved from <http://www.apa.org/pubs/journals/amp/>
- Schwarz, N. & Oyserman, D. (2001). Asking questions about behavior: Cognition, communication, and questionnaire construction. *American Journal of Evaluation*, *22*(2), p. 127-160.
doi:10.1177/109821400102200202
- Sibthorp, J., Paisley, K., Gookin, J., & Ward, P. (2007). Addressing response-shift bias: Retrospective pretests in recreation research and evaluation. *Journal of Leisure Research*, *39*(2), p. 295-315.
- Sprangers, M., & Hoogstraten, J. (1989). Pretesting effects in retrospective pretest-posttest designs. *Journal of Applied Psychology*, *74*(2), 265-272. Retrieved from <http://www.apa.org/pubs/journals/apl/>
- Taylor, P. J., Russ-Eft, D. F., & Taylor, H. (2009) Gilding the outcome by tarnishing the past: Inflationary biases in retrospective pretest. *American Journal of Evaluation*, *30*(1), p. 31-34.
doi:10.1177/1098214008328517