
Evidence to Support the Use of the Retrospective Pretest method to Measure Dietary and Physical Activity Behavior and Self-Efficacy in Adolescents

Mical Kay Shilts

Department of Family and Consumer Sciences
California State University Sacramento
shiltsm@csus.edu

Dorothy Smith

University of California Cooperative Extension
dorsmith@ucdavis.edu

Lenna Ontai

Department of Human and Community Development
University of California Davis
lontagr@ucdavis.edu

Marilyn S. Townsend

Department of Nutrition
University of California Davis
mstownsend@ucdavis.edu



Evidence to Support the Use of the Retrospective Pretest method to Measure Dietary and Physical Activity Behavior and Self-Efficacy in Adolescents

Mical Kay Shilts
California State University Sacramento

Dorothy Smith, Lenna Ontai and Marilyn S. Townsend
University of California Davis

Abstract: This study compared the retrospective pretest-posttest method to the traditional prospective pretest-posttest method assessing adolescents' dietary and physical activity, self-efficacy and behaviors. Participants were 7th and 8th grade students at a rural K-8th grade elementary school in Northern California (n=188). All participants completed an evaluation instrument (traditional pretest), followed by a 9-lesson dietary and physical activity intervention. Upon completion of the intervention, participants completed a second and identical evaluation instrument which served as a traditional posttest. The following day, participants completed another evaluation tool, this time formatted as a retrospective pretest. Analysis included sample *t* tests comparing the means of each method. Participants (n = 154) with a mean age of 13 ± .7 years old were included in the analyses (52% female). Paired sample *t* tests reported non-significant differences between the two methods for dietary behavior and dietary self-efficacy, yet significant differences were found for physical activity behavior ($p < .05$) and physical activity self-efficacy ($p < .01$). We conclude that the retrospective pretest-posttest method was as good a measure of dietary self-efficacy and behavior as the traditional prospective pretest-posttest method and may be better at attenuating response-shift bias when assessing physical activity self-efficacy and behavior.

Introduction and Background

Evaluation is an important component of Youth Development programmatic efforts designed to change knowledge and behaviors. A one-group pretest-posttest design is often selected, where the evaluator uses a traditional prospective pretest-posttest method of data collection (Campbell & Stanley, 1966; Smith, Genry, & Ketring, 2005). This method has limitations in real world application, especially with adolescents. Administering tests on two separate occasions may result in additional burden and uses vital program delivery time for the Extension educator (Raidl et al., 2004).

When working with adolescents, establishing rapport at the first educational meeting is important for learning. Test taking can be perceived as intrusive and an obstacle to establishing trust (Stipek, 2002). Administering a pretest at the beginning of the program could reduce pretest completion rates because not all who participate in the program are in attendance at the pretest date (Raidl et al., 2004). In addition to practical barriers, the traditional prospective pretest-posttest method can result in response-shift bias where the participant rates himself differently on the posttest after acquiring new information related to the test item (Cook & Campbell, 1979; Howard & Dailey, 1979). The standard of measurement shifts from the first administration of the test (pretest) to the second (posttest) (Cook & Campbell, 1979; Howard & Dailey, 1979). The result is an underestimated effect during program evaluation (Hill & Betz, 2005). Such underestimation may be prevalent in nutrition-related programs where adolescents are likely to overestimate their initial capabilities. After the program, they may be more realistic about their capabilities, masking actual changes with a prospective pretest-posttest (Howard et al., 1979).

Using a retrospective pretest-posttest method instead of a traditional prospective pretest-posttest method could be a viable substitute for youth program evaluation. This method asks youths to take one test at the end of the program responding to each evaluation question twice. First, respondents answer each item in a traditional posttest manner, reporting current behaviors or skills. Then they respond to the same items retrospectively, before the program began.

The retrospective pretest-posttest method has been recommended when conducting program evaluation using self-report measures because the design encourages participants to rate themselves from the same perspective producing a more legitimate evaluation of program outcomes (Aiken & West, 1990; Howard & Dailey, 1979; Howard et al., 1979; Pratt, McGuigan, & Katzeva, 2001; Rohs, Langone, & Coleman, 2001). Evaluation of the retrospective pretest-posttest method resulted in positive results with adults addressing various behaviors such as parenting (Pratt et al., 2001), nutrition and resource management (Raidl et al., 2004; Rockwell & Kohn, 1989), food safety and menu planning (Rohs et al., 2001). The retrospective pretest-posttest method may be a promising avenue for youth program evaluations, especially for program content where overestimation of initial capabilities are likely.

Two major literature reviews of dietary and physical activity evaluation tools for low-income youth and adults discuss the importance and need for low respondent burden measures for these audiences (Contento, Randell, & Basch, 2002; McClelland et al., 2001; Townsend, 2006). While the retrospective pretest method may be less burdensome, we found no studies comparing this method relative to the traditional pretest-posttest method with youth for dietary and physical activity behaviors and self-efficacy. Given the importance of evaluation and federal reporting for Extension programs, it is valuable to determine if the retrospective pretest method is viable with adolescents.

Purpose

The purpose of this study was to investigate the retrospective pretest-posttest method as an appropriate alternative to the traditional prospective pretest-posttest for evaluation of a youth development intervention with adolescents.

Methods and Procedures

Procedure

All participants completed an evaluation instrument (traditional prospective pretest) assessing dietary and physical activity behaviors and self-efficacy, followed by a 9-lesson dietary and physical activity intervention, called EatFit (Horowitz, Shilts, & Townsend, 2004). Upon completion of the intervention, participants completed a second and identical evaluation instrument which served as a traditional posttest. The following day, participants completed another evaluation tool, this time formatted as a retrospective pretest. This version was similar to the pre and posttests but the question headings read, "These questions are asking about your food/physical activity choices before EatFit. Think back 6 weeks, before you had any EatFit lessons. Fill in the bubble that best describes your food/physical activity choices before the EatFit classes began." The Extension staff person teaching the intervention was a community nutrition educator for California Youth Food Stamp Nutrition Education (FSNE) and was trained and certified to teach the intervention. The study protocol was approved by the Institutional Review Board at the University of California, Davis.

Sample

The sample was drawn from a low-income ($\geq 50\%$ qualified for free or reduced price school lunch), rural K-8th grade elementary school in Northern California. The participants were all 7th and 8th grade students ($n=188$) participating in physical education. Of the 188 potential participants, 28 participants did not return both consent and assent forms and 6 did not complete the evaluation instruments leaving a total number of 154 participants.

Measures

A self-administered instrument assessed participants' dietary behaviors (19 items), physical activity behaviors (4 items), dietary self-efficacy (19 items), and physical activity self-efficacy (4 items). The items in the dietary and physical activity behavior sections were adapted from the Centers for Disease Control Youth Risk Behavior Survey (YRBS) (Center for Disease Control and Prevention, 2007). The YRBS dietary and physical activity items were modified slightly to include specific targeted behaviors of the intervention. Self-efficacy was defined as confidence to perform a targeted behavior and was determined by asking participants to self-report confidence to perform targeted behaviors (Bandura, 1986). Response range for the behavior-related items was an 8-point scale signifying the number of days per week the participant engaged in the targeted behavior, i.e., 0-7 days per week. The response range for the self-efficacy items was a 4-point scale, i.e., 1 equaling not at all confident to 4 being totally confident.

Reliability testing of the Youth Risk Behavior Survey (YRBS) items with a nationally representative sample of adolescents indicated Kappas ranging from 91.1-64.2% (Brener, Collins, Kann, Warren, & Williams, 1995). Using the concurrent method of Willis (1994), all items adapted for this study were cognitively tested (Contento et al., 2002). In individual interviews with 8th grade students ($n=16$), items were revised and retested using 4 questions:

- *What does the question mean to you using your own words?*
- *How did you come up with your answer?*

- *Think about other students in your grade at school. Are there any difficult words for them?*
- *How would you make this question clearer to them?*

Items were evaluated for content validity by three experts in behavioral nutrition and found to represent the corresponding construct domain. The instrument was pilot tested with 6-8th graders (n=34) (Shilts, Townsend, & Horowitz, 2002).

A reliability assessment of the revised instrument was conducted to establish that the items were measuring phenomena in a reproducible and consistent way (Carmines & Zeller, 1979; Litwin, 1995). Seventh and 8th grade students (n=46) completed the instrument on two occasions, three weeks apart, with no intervention. Reliability coefficients were .73 for the dietary behavior items, .55 for the physical activity behavior items, .59 for dietary self-efficacy items and .48 for physical activity self-efficacy items. Scales and instruments used with adults are thought to have good test-retest reliability with coefficients of .7 or greater (Litwin, 1995). The coefficients for the dietary behavior items met this criterion. The other coefficients are lower than .7, indicating more random error associated with the items. Because this reliability assessment was conducted with 12-14 year olds, we are considering them marginally acceptable for our purposes. The retrospective version of the evaluation instrument was additionally cognitively tested with adolescents using the same methods as previously described and then revised (n=12).

Intervention

The youth development intervention is a National 4-H juried curriculum called EatFit and is designed to improve the dietary and physical activity behaviors of adolescents 11-14 years old (Horowitz et al., 2004; Shilts, Townsend, & Horowitz, 2004). The intervention includes nine experiential lessons driven by Social Cognitive Theory (SCT) and uses web-based assessment (www.eatfit.net) to assist participants in dietary analysis and goal setting (Horowitz, Shilts, & Townsend, 2005). A description of the intervention has been reported previously (Horowitz et al., 2004; Shilts et al., 2004). This intervention was designed specifically for three U.S. Department of Agriculture youth programs in California: EFNEP, FSNE and 4-H.

Analysis

Statistical analyses were conducted using SAS PC version 8.1 (SAS Institute, Inc). Double data entry in two separate files was performed and each file was compared for differences using the *compare procedure*. Paired sample *t* tests compared the difference in mean scores between the traditional pretest and the retrospective pretest for each variable, as well as between the retrospective pretest and posttest and the traditional pretest and posttest. Pearson correlation was used to explore the relationship between the two pretest methods to investigate if they were comparable measures.

Results

The mean age of the participants was 13 ± .7. More than half (52%) were female. Participants self-reported as non-Hispanic white (69%), Hispanic (14%), multi-ethnic (14%), Asian/Pacific Islander (2%), and American Indian (2%) [Table 1].

Table 1

Demographic Characteristics of Study Participants from
A Rural Middle School in California (n=154)

| Gender | | n | % |
|------------------|-----------------|----------|----------|
| | Male | 74 | 48.1 |
| | Female | 80 | 51.9 |
| Ethnicity | | | |
| | American Indian | 3 | 1.9 |
| | Asian | 3 | 1.9 |
| | White | 106 | 68.8 |
| | Hispanic | 21 | 13.6 |
| | Multi Ethnic | 21 | 13.6 |
| Age | | | |
| | 12 | 39 | 25.3 |
| | 13 | 80 | 51.9 |
| | 14 | 34 | 22.1 |
| | 15 | 1 | 0.6 |
| Grade | | | |
| | 7 th | 68 | 44.2 |
| | 8 th | 86 | 55.8 |

Mean scores for each test (retrospective pre, traditional prospective pre, post) and differences are reported in Table 2. Paired sample *t* tests reported non-significant differences between the two methods (retrospective and traditional prospective) for dietary behavior ($p = .58$) and dietary self-efficacy ($p = .22$), yet significant differences were found for physical activity behavior ($p = .02$) and physical activity self-efficacy ($p = .002$) [Table 2].

Table 2

Traditional Prospective Pretest (Pre), Retrospective Pretest (Retro),
And Posttest (Post) Mean Scores and Paired *t*-test Results for the Differences (n=154)

| Variable | Pre | Retro | Post | ←----- Differences -----> | | |
|---------------------------------|--------------|--------------|-------------|-------------------------------------|------------------|--------------------|
| | | | | Retro- Pre | Pre- Post | Retro- Post |
| Dietary Behavior | 63.42 ±15.19 | 62.76±13.85 | 67.49±13.91 | -0.66±14.60 | 4.07±12.70*** | 4.77±13.75**** |
| Physical Activity Behavior | 19.82±5.63 | 18.68±5.53 | 18.64±6.21 | -1.14±5.87* | -1.17±5.91* | -0.04±5.49 |
| Dietary Self-Efficacy | 54.90±9.72 | 53.94±10.14 | 57.25±9.99 | -0.95±9.71 | 2.29±7.85*** | 3.44±8.45**** |
| Physical Activity Self-Efficacy | 13.97±2.40 | 13.37±2.55 | 14.20±2.31 | -0.68±2.57** | 0.16±1.73 | 0.91±2.45**** |

* $p < .05$, ** $p < .01$, *** $p < .001$, **** $p < .0001$

The correlations (Pearson correlation coefficient) between the retrospective pretest and the traditional prospective pretest for each variable were: dietary behavior ($r = .50$, $p < .0001$), dietary self-efficacy ($r = .52$, $p < .0001$), physical activity behavior ($r = .44$, $p < .0001$) and physical activity self-efficacy ($r = .57$, $p < .0001$). These results indicate that the pretests were moderately correlated.

Discussion

Our results suggest that the retrospective pretest method was as good a measure of dietary self-efficacy and behavior in adolescents as the traditional pretest method, and may be better at attenuating response-shift bias when assessing physical activity self-efficacy and behavior.

Correlations between the retrospective pretest and the traditional prospective pretest investigated if participants had reliable recall retrospectively relative to the prospective method. To illustrate that the retrospective pretest was as good as the traditional prospective pretest, the test scores would have the same mean (paired t test) and would be correlated. The paired t -test revealed no differences between the retrospective pretest and the traditional pretest for dietary self-efficacy and behavior variables and they were significantly correlated. The results suggest that the retrospective pretest for the dietary variables was comparable to the traditional pretest and could be substituted when appropriate.

Bias is inherent in all self-report instruments regardless of how administered (Hill & Betz, 2005). Bias is systematic error that reduces the validity of an evaluation tool. The traditional prospective pretest may encounter a response-shift related to participants' changed standards of measurement after the intervention (Rohs et al., 2001). With the retrospective pretest-posttest method, the standard of measurement is the same for both tests. However, the retrospective format is vulnerable to other threats to validity such as motivational bias (Hill & Betz, 2005). The ultimate goal of a well-developed evaluation tool is to minimize the sources of bias, while addressing administration and client burden issues.

A common issue with the retrospective pretest method with adults is overestimation of program effect caused by motivational factors such as justification of effort or socially desirable change (Hill & Betz, 2005). Motivational bias may influence the outcome of a lifestyle intervention such as EatFit. When adolescents have spent time and effort participating in EatFit and then are asked to rate their physical activity change or improvement from before to after the program, behavior and self-efficacy change scores may be inflated as a result of this particular bias (Hill & Betz, 2005). However, this type of bias could affect both testing formats with an inflated posttest. Our results suggest that this type of bias was not present for the dietary behavior and self-efficacy measures.

The pattern of results for physical activity was different from the pattern for dietary items. In comparison to the dietary variables, the physical activity variables were significantly correlated, but the paired t tests indicated different means, with students scoring themselves lower on the retrospective pretest. This difference may indicate that physical activity is particularly vulnerable to response-shift bias found in other traditional pre-post test studies of adults resulting in an underestimation of the program results. This interpretation seems reasonable given that the traditional prospective pretest results showed students decreasing physical activity behaviors while making no change in self-efficacy after receiving nine lessons of a nutrition and physical activity intervention. In addition, the results from the prospective pretest method do not support the Social Cognitive Theory's (SCT) construct where change in self-efficacy precedes change in behavior (Bandura, 1986). However, the retrospective pattern of results is consistent with SCT, with a statistically significant increase in physical activity self-efficacy but not for physical activity behaviors. The correlation between change in physical activity behavior and change in physical activity self-efficacy using the retrospective pretest was significant ($r = .41, p < .0001$), indicating that behavior change was related to change in self-efficacy as anticipated by SCT. The same was true for the relationship between change in dietary behavior and change in self-efficacy ($r = .50, p < .0001$).

Although the students made significant improvements in dietary behaviors after receiving the EatFit intervention, they did not do so for physical activity behaviors. It is possible that the nature of the intervention impeded the impact of the program since the intervention took place during physical education class. The EatFit lessons were substituted for the participants' normal physical activity which may have contributed to a decrease in their ratings of physical activity level. In addition, the time period of the intervention occurred during an extremely rainy season of the year, increasing barriers to physical activity.

Limitations

A major weakness of our study was that there was no gold standard against which to compare the two test formats. Observational data would serve such a role. Future research should incorporate sources of objective criteria with which to compare both types of traditional prospective and retrospective self-report test data.

Conclusion

Our results indicate that the retrospective pretest-posttest method of data collection is a viable substitute for the traditional prospective pretest-posttest method to measure dietary and physical activity self-efficacy and behaviors in adolescents.

Implications

Quality and meaningful learning is the goal of all youth development programs and activities such as 4-H, after school programs, Youth FSNE and Youth EFNEP. The traditional prospective pretest-posttest method of evaluation has probably been the most utilized but not without limitations. The ability to reduce the testing burden on the youth participant while maintaining the ability to assess knowledge or attitude change is extremely valuable. By reducing the "test" to a one time event 4-H staff might be able to increase data collection from volunteer leaders. Increasing data collection opportunities will allow staff to better evaluate a larger variety of programs and projects being offered and may also provide insight into volunteer leader's skills and training needs.

The use of the retrospective pretest would work well for all types of Youth programs including:

- Junior Master Gardener
- Wildlife Habitat Evaluation Program
- Environmental Education
- Shooting Sports
- Animal Science - Large Animal and Small Animal
- Engineering and Technology
- Family and Consumer Science
- Plant Science
- Health and Leisure
- Social Science - Personal Development and Resource Science.

In addition, the study indicated this method was a good measure for the EatFit intervention and may be applicable to other nutrition programs used in Extension such as Power Play and Jump Start. By reducing the testing to a one time event, the learning time can actually be increased. This is a valuable marketing strategy for staff trying to promote these programs to teachers.

Other Cooperative Extension programs would benefit from these findings. The data collected from an evaluation can provide valuable information for program improvement and enhancement. In addition, evaluation is essential when applying for continued support. The retrospective pretest-posttest method is an appropriate method for an organization to get an accurate picture of participants' perceptions of benefit in addition to participant reports of skills and behaviors. Participants can complete the evaluation in one relatively short session; this can be beneficial when working with any youth group including camps, sports, scouts, etc... In these non-formal educational settings, this method might be more readily accepted by participants as it is less burdensome and "academic" as compared with the traditional pretest-posttest method.

References

- Aiken, L.S., & West, S.G. (1990). Invalidation of true experiments self-report pretest biases. *Evaluation Review, 14*(4), 374-390.
- Bandura, A. (1986). *Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Brener, N.D., Collins, J.L., Kann, L., Warren, C.W., & Williams, B.I. (1995). Reliability of the Youth Risk Behavior Survey questionnaire. *American Journal of Epidemiology, 141*(6), 575-580.
- Campbell, D.T., & Stanley, J.C. (1966). *Experimental and Quasi-experimental Designs for Research*. Chicago: Rand McNally.
- Carmines, E., & Zeller, R. (1979). *Reliability and Validity Assessment*. Newbury Park, CA: Sage Publications.
- Center for Disease Control and Prevention (2007). *Youth Risk Behavior Survey*. Accessed: July 2, 2007. Available online at: <http://www.cdc.gov/HealthyYouth/yrbs/pdf/questionnaire/2007HighSchool.pdf>
- Contento, I.R., Randell, J.S., & Basch, C.E. (2002). Review and analysis of evaluation measures used in nutrition education intervention research. *Journal of Nutrition Education and Behavior, 34*(1), 2-25.
- Cook, T.D., & Campbell, D.T. (1979). *Quasi-Experimentation: Design and Analysis for Field Settings*. Chicago: Rand McNally.
- Hill, L.G., & Betz, D.L. (2005). Revisiting the retrospective pretest. *American Journal of Extension, 26*(4), 501-517.
- Horowitz, M., Shilts, M.K., & Townsend, M.S. (2004). EatFit: a goal oriented intervention that challenges middle school adolescents to improve their eating and fitness choices. *Journal of Nutrition Education and Behavior, 36*(1), 43-44.
- Horowitz, M., Shilts, M.K., & Townsend, M.S. (2005). Adapting a diet analysis program for an adolescent audience. *Journal of Nutrition Education and Behavior, 37*, 43-44.

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.

Howard, G.S., Ralph, K.M., Gulanich, N.A., Maxwell, S.E., Nance, D., & Gerber, S.L. (1979). Internal invalidity in pretest-posttest self-report evaluations and the re-evaluation of retrospective pretests. *Applied Psychological Measurement, 3*, 1623.

Litwin, M.S. (1995). *How to Measure Survey Reliability and Validity*. Thousand Oaks: Sage Publications.

McClelland, J.W., Keenan, D.P., Lewis, J., Foerster, S., Sugerman, S., Mara, P., et al. (2001). Review of evaluation tools used to assess the impact of nutrition education on dietary intake and quality, weight management practices, and physical activity of low-income audiences. *Journal of Nutrition Education, 33*(S1), S35-S48.

Pratt, C., McGuigan, W., & Katzeva, A. (2001). Measuring program outcomes: using retrospective pretest methodology. *American Journal of Evaluation, 21*, 341-349.

Raidl, M., Johnson, S., Gardiner, K., Denham, M., Spain, K., Lanting, R., et al. (2004). Use retrospective surveys to obtain complete data sets and measure impact in extension programs. *Journal of Extension, 42*(2). Available online at: <http://www.joe.org/joe/2004april/rb2002.shtml>

Rockwell, S.K., & Kohn, H. (1989). Post-then-pre evaluation. *Journal of Extension, 27*(2). Available online at: <http://www.joe.org/joe/1989summer/a1985.html>

Rohs, F.R., Langone, C.A., & Coleman, R.K. (2001). Response shift bias: a problem in evaluating nutrition training using self-report measures. *Journal of Nutrition Education, 33*(3), 165-170.

SAS Institute, Inc. (Version 8.1). Cary, NC.

Shilts, M.K., Townsend, M.S., & Horowitz, M. (2002). Pilot study of the EatFit intervention to determine sample size and protocol for a randomized controlled trial. *Center for Advanced Studies in Nutrition and Social Marketing, University of California, Davis*. Available online at: <http://socialmarketing-nutrition.ucdavis.edu/publications.htm#review6>

Shilts, M.K., Townsend, M.S., & Horowitz, M. (2004). An innovative approach to goal setting for adolescents: guided goal setting. *Journal of Nutrition Education and Behavior, 36*(3), 155-156.

Smith, T.A., Genry, L.S., & Ketring, S.A. (2005). Evaluating a Youth Leadership Life Skills Development Program. *Journal of Extension, 43*(2), Available online at: <http://www.joe.org/joe/2005april/rb2003.shtml>

Stipek, D. (2002). *Motivation to Learn: Integrating Theory and Practice*. 4th ed. Boston: Allyn & Bacon.

Townsend, M.S. (2006). Evaluating food stamp nutrition education: process for development and validation of evaluation measures. *Journal of Nutrition Education and Behavior, 38*, 18-24.

Willis, G. (1994). *Cognitive Interviewing and Questionnaire Design: A Training Manual (Working Paper Series No. 7)*. Hyattsville, MD: Centers for Disease Control and Prevention, National Center for Health Statistics.

Acknowledgements

We wish to thank the Food Stamp Nutrition Education and UCCE for funding this project. We also wish to thank Jan Peerson, Calaveras County Youth FSNE staff, and Mr. Keller and Mr. Owen and their students for making this study possible.