DOCUMENT RESUME

ED 286 896 TM 870 518

AUTHOR Sprangers, Mirjam
TITLE Validity Threats in Retrospective Pretest-Posttest Designs.
PUB DATE 24 Apr 87
NOTE 13p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, DC, April 20-24, 1987).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Higher Education; *Pretests Posttests; Program Evaluation; Research Design; *Response Style (Tests); *Self Concept Measures; *Self Evaluation (Individuals); Social Desirability; Testing Problems; *Test Validity
IDENTIFIERS Bogus Pipeline Induction; *Response Shift; *Retrospective Pretesting

ABSTRACT

Social desirability responding and effort justification were each studied for their effect on response shift as seen in retrospective pretest-posttest responses. The first study used a bogus pipeline induction (half the subjects were told their self-reporting could later be validated) before the self-report pretest. After experimental or control treatment they completed a 16-scale self-reporting instrument. The treatment was found to be effective, with bogus pipeline induction lowering self-reported preratings and preventing response shift. The second experiment examined effort justification, lowering the pretest rating to justify participation effort. It included experimental, placebo, and no-treatment conditions. Both a self-report instrument consisting of 20 self-rating scales and a 16 three-choice item knowledge test were administered. The treatment was found to be effective and a mean pre-retrospective difference was found in both experimental and placebo conditions. Pre- and retrospective self-ratings in no-treatment controls did not differ significantly. Since response shift is a severe validity threat in conventional pretest-posttest designs using self-report measures, the retrospective pretest was recommended to control for this confounding factor. If effort justification is controlled for and social desirability can be ruled out, response shift may not occur. (MGD)

Validity Threats in Retrospective Pretest-Posttest Designs

Paper presented at the AERA Division D program:
Problems in Evaluation Research Designs

Washington D.C., April 24th 1987

Mirjam Sprangers
Department of Psychology
University of Amsterdam
Weesperplein 8
1018 XA Amsterdam
the Netherlands

Response shift

If a treatment aims to alter participants' understanding of the target concept, subjects may change their internal standard as a result of the training. Howard, Ralph, Gulanick, Maxwell, Nance, & Gerber (1979) identified this response shift as a potential confounding influence when evaluating results from pretest-posttest designs that employ self-reports as outcome criteria. Since a response shift renders pre and posttest scores incompatible, pretest-posttest comparisons within the experimental condition are invalidated. Posttest comparisons between the experimental and control condition are confounded as well.

Retrospective pretest

Howard and his colleagues recommended the use of retrospective ratings to control for response shift bias effects. After completing the posttest conventionally, subjects keep the posttest in front of them and are then asked to report how they now perceive themselves to have been prior to the training. Subjects react to each retrospective item in relation to the answer given to the corresponding posttest item. (I might add that this procedure in administering the tests is different from that of Howard and his colleagues who followed an item-after-item procedure). It was hypothesized that the posttest and retrospective pretest would be filled out with respect to the same internal standard. Consequently, comparison of the posttest and retrospective pretest scores would eliminate a treatment produced response shift. By now, a considerable amount of studies performed within an educational training context, favored the retrospective pre-post difference scores, in providing a more accurate estimate of a treatment effect, while the conventional pre-post difference scores most often masked the treatment effect.

To illustrate, when a communication skills training is evaluated, a self-report item may be like: *"In a heated discussion I am still able to listen to what others are saying"*. Subjects can rate their responses on a scale ranging from 1, not at all applicable to me, to

7, completely applicable to me. During the **pretest** a subject may say: *"I listen to what other people say when I'm talking to them, I'd say 6"* (very applicable to me). However after the training the subject may say: *"All these group excercises made me realise that I don't listen to people. I should have put 1 the first time I filled this out"*. So **retrospectively**, the pretest item should be filled out a 1 (not at all applicable to me). At the **posttest** this subject may say: *"The group really opened my eyes and helped me to try to be more of an active listener and so while I still sometimes forget to listen to people, overall I'm not doing so badly now. I'll put a 5"* (rather applicable to me) (Howard, et al., 1979, pp. 3-5). In this example a comparison of conventional pre and postratings would show a decline of 6 to 5, whereas the comparison of retrospective pre and postratings would indicate a positive treatment effect (1 to 5).

Validity threats

Although the retrospective pretest-posttest design controls for response shift bias effects, it is susceptible to a variety of other validity threats. The measurement of the central concept 'response shift' poses serious problems, in particular. The phenomenon is operationalized on group level, by the mean difference between self-reported pre and retrospective preratings. When a substantial mean difference is found in the experimental condition and the difference is negligible in the control condition, a response shift is claimed to have occurred. However, a serious complication arises from the fact that several alternative explanations may also account for a mean difference between conventional and retrospective preratings. Two such alternative explanations are social desirabilty responding and effort justification. The purpose of the present paper is to discuss these two confounders in light of the results of two experiments, performed within an educational training context.

i. Social desirability responding

It is possible that retrospective ratings, rather than representing the true level of functioning, represent impression management.

The first experiment investigated the operation of social desirability. After Howard, Millham, Slaten, & O'Donnell (1981), a prevention or reduction of social desirability was planned, utilizing a bogus pipeline technique. Half of the subjects were led to believe that the veracity of their self-reports could be checked by means of objective measures. The experimental design is displayed in Table 1. **Subjects** were psychology freshmen of the University of Amsterdam, who were fulfilling a course requirement. The **experimental treatment** consisted of a programmed instruction on Seeing Problems Strategies; subjects were instructed and trained in producing suggestions for the improvement of common appliances. This treatment took about one hour. The **control treatment** consisted of summaries of research procedures; subjects were asked to give their opinion on the ethical permissibleness. The control treatment was of similar format and lay-out as the experimental treatment and took the same amount of time. The **self-report** instrument consisted of 16 7-point scales about the topics trained. To illustrate: *"I would be good in suggesting methods for preventing bicycle theft"* and *"I don't believe I am so inventive in improving common appliances"*. The **objective measure** to assess subjects' performance, consisted of 10 common appliances (different from those used in training) for which subjects had to produce suggestions for improvement. The **bogus pipeline induction** took place prior to the administration of the self-report pretest in experimental conditions 1 and 2, and control conditions 5 and 6. A written announcement was made that in the course of the experimental session the truthfulness of the self-reports was to be verified. This announcement was repeated on the self-report post and retrospective pretest. Both pre and posttreatment performance meausures were administered to induce the bogus pipeline deception. In addition, we investigated the robustness of the retrospective pretest to procedural differences. We therefore reversed

the order of the posttest and retrospective pretest in conditions 2, 4 and 6, and examined whether administration of the retrospective pretest, independent of the posttest affects the ratings.

Results are as follows. 1. The treatment was effective, with regard to both performance and self-report measures. Experimental subjects performed significantly better than controls.

2. The retrospective pretest is rather robust for order manipulations; neither postratings nor retrospective ratings were affected. We therefore combined conditions 1 and 2, 3 and 4, and 5 and 6 respectively.

3. The bogus pipeline induction did lower self-reported pre ratings and prevented the occurrence of a response shift. Thus, the only significant conventional pre-retrospective pre difference took place in combined conditions 3 and 4 (see Figure 1). The conclusion must be that social desirability responding is a viable alternative explanation of the response shift phenomenon.

On first sight, this result is in contrast with that of Howard, Millham, Slaten & O'Donnell (1981), who still found pre-retrospective differences under bogus pipeline conditions. However, they exposed their subjects to the bogus pipeline induction at the posttesting. Indeed, retrospective scores were not affected by the bogus pipeline induction in our experiment either. However, we consider a bogus pipeline induction at pretesting preferable and more informative, since a response shift is defined by a mean difference between conventional pre and retrospective pre scores and, moreover, a pre-retrospective difference can be caused by an initial over- or underrating due to social desirability responding,

2. Effort justification

When subjects do not experience any benefit of the training, they may, in an attempt to justify the effort spent, adjust their initial pre-treatment ratings in a downward direction.

The second study made use of a design that incorporated a placebo control condition, in addition to an experimental and a no-treatment control condition. Since placebo subjects devote the same amount of time and effort to the placebo treatment as do experimental subjects to the experimental training, a mean difference in pre-retrospective self-ratings in the placebo condition can be explained in terms of effort justification, thus invalidating the response shift interpretation. The experimental design is diagrammed in Table 2. Again, **subjects** were psychology freshmen who participated in the experiment in exchange for course credit. To summarize, the **experimental treatment** consisted of a film about childrens' play activities. The film took 25 minutes and followed a programmed instruction procedure. The **placebo treatment** consisted of a film on communication skills. The film followed a similar procedure and took the same amount of time as the experimental film. During the experimental and placebo interventions, **no-treatment** control subjects were sent away. The **self-report** instrument consisted of 20 7-point scales. To illustrate: *"I know what kind of play activities are common for children from 2 to 5 years"*. The **objective measure** was a knowledge test, consisting of 16 three-choice items.

The results are: 1. The treatment was effective; both performance and self-report indices of change reached significance in the experimental condition only.

2. A mean pre-retrospective difference was found in the experimental condition. However, in the placebo condition a significant difference was found too. Conventional pre and retrospective self-ratings of the no-treatment control subjects did not differ significantly (see Figure 2). Since the placebo treatment did produce lower retrospective ratings, results lend support for the hypothesis of effort justification.

Conclusions

On conclusion, response shift represents a severe validity threat to conventional pretest-posttest designs that employ self-report measures as outcome criteria. Researchers who evaluate educational trainings shculd be aware of the potential occurrence of a response shift. The retro ective pretest seems to be a potentially useful extension of conventional pretest-posttest designs in controlling for response shift bias. However, methodological problems associated with the retrospective pretest-posttest design are real and probably underestimated when results of the studies published sofar, are taken into account only. For, to recapitulate the present outcomes: a response shift may not occur if 1) effort justification is controlled for, and 2) social desirability can be ruled out. I want to stress though, that the actual occurrence of these confounding influences depends on the specific experimental setting, the nature of the intervention and the corresponding measures. Researchers should be aware of these potential confounders of the response shift interpretation and they should design their experiments accordingly. In addition, they must keep in mind that the alternative of the conventional pretest-posttest design may be even more vulnerable to validity threats.

References

Howard, G.S., Millham, J., Slaten, S., & O'Donnell, L., 1981. Influence of subject response style effects on retrospective measures. *Applied Psychological Measurement, 5*. 89-100.

Howard, G.S., Ralph, K.M., Gulanick, N.A., Maxwell, S.E., Nance, D W., & Gerber, S.K., 1979. Internal invalidity in pretest-posttest self-report evaluations and a re-evaluation of retrospective pretests. *Applied Psychological Measurement. 3*. 1-23.

Sprangers, M., & Hoogstraten, Joh., 1987. Response shift bias and a bogus pipeline; two experiments. Submitted.

Sprangers, M., & Hoogstraten, Joh., 1987. On delay and reassessment of retrospective ratings. Submitted.

Sprangers, M., & Hoogstraten, Joh., 1987. Pretesting effects in retrospective pretest-posttest designs. Submitted.

Sprangers, M., & Hoogstraten, Joh., 1987. Effort justification in retrospective, improvement and recall ratings. Submitted.

Table 1: Experimental design

condition	n	bogus pipeline induction	self-report pretest	objective pretest	treatment	self-report	objective posttest
1	12	yes	yes	yes	experim.	post-retro	yes
2	12	yes	yes	yes	experim.	retro-post	yes
3	13	no	yes	yes	experim.	post-retro	yes
4	12	no	yes	yes	experim.	retro-post	yes
5	12	yes	yes	yes	control	post-retro	yes
6	12	yes	yes	yes	control	retro-post	yes

Figure 1: graphic presentation of mean self-reported pre, post and retrospective prescores

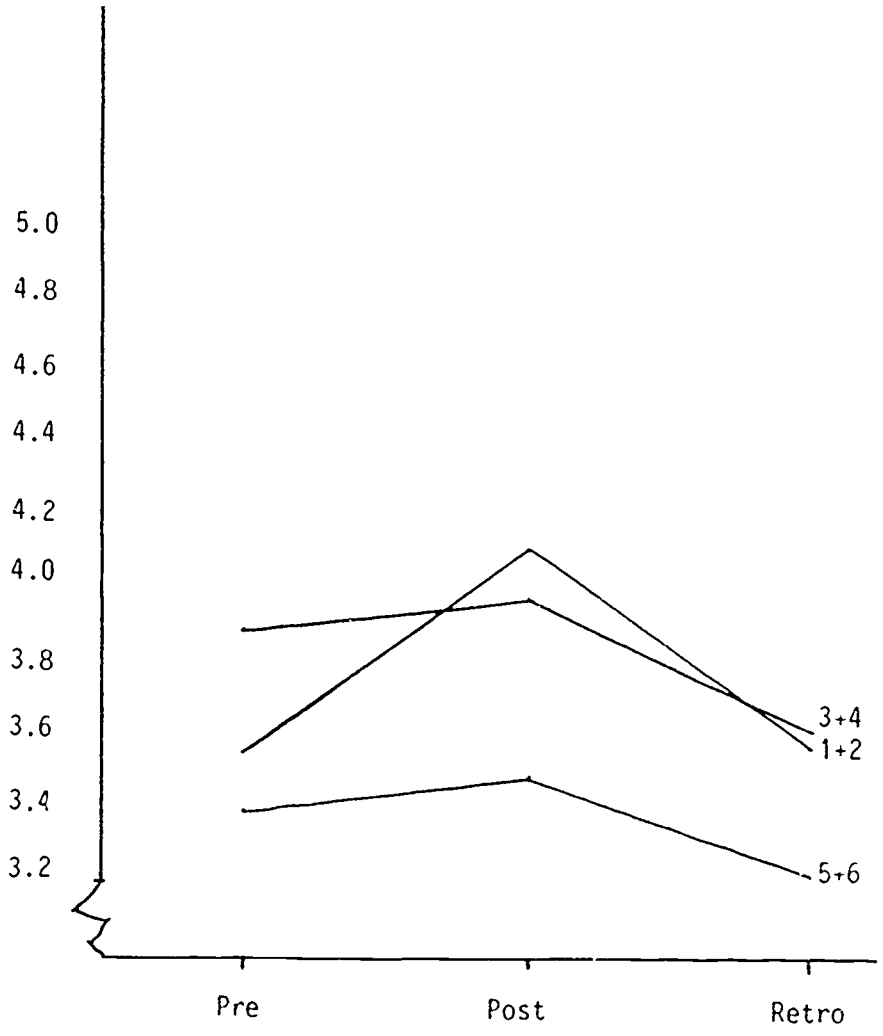

Table 2: Experimental design

condi-tion	n	self-report pre	object. pre	treatment	self-report post-retro	object. post
1	17	yes	yes	experimental	yes	yes
2	14	yes	yes	placebo	yes	yes
3	15	yes	yes	control	yes	yes

Figure 2: graphic presentation of mean self-reported pre, post and retrospective prescores