Bulletin of the Psychonomic Society
1988, 26 (2), 124-125

A method for analyzing retrospective pretest/posttest designs: II. Application

JOHAN HOOGSTRATEN and PIETER KOELE
University of Amsterdam, Amsterdam, The Netherlands

Koele and Hoogstraten (1988) argued that, when using a retrospective pretest/posttest design, a test on the presence of a response-shift bias is necessary before reliable estimates of a treatment effect can be obtained. To that purpose, they developed a hierarchical model fitting procedure aimed at detecting the presence of a treatment effect and/or a response-shift bias. In the present article this procedure was applied in a reanalysis of the results of an earlier study.

In retrospective pretest/posttest designs, a treatment effect can be estimated by using either pretest/posttest difference scores or posttest/thentest difference scores. If a response shift has occurred, as indicated by a substantial mean difference between scores on the conventional pretest and the retrospective pretest (or thentest), posttest/thentest difference scores give a reliable estimate of the treatment effect; if there is no response-shift bias, pretest/posttest difference scores should be used for estimating the treatment effect. Koele and Hoogstraten (1988) developed a method for detecting the presence of a response-shift bias. They recommended that this method be used to decide which difference scores should be used for estimating the treatment effect. In this article the application of their method will be elucidated by reanalyzing the results of an experiment by Hoogstraten (1982).

In Hoogstraten's (1982) study the relative validity of the retrospective pretest/posttest design was investigated in an educational training context. For our purposes, two of the three conditions used in the experiment are relevant. In both these conditions all subjects filled out pre-, post-, and thentests (self-reports on their level of ability). The tests consisted of 17 items with 7-point rating scales. Test scores were taken to be the sum of the item scores. In the experimental condition (n = 25), subjects received specific training between pre- and posttest; in the control condition (n = 21), subjects received no such training. It was expected that the training would have a beneficial effect on the test scores and that a potential response-shift bias in the experimental condition would lead to lower scores on the thentest than on the pretest. On the basis of these specifications, the data will be analyzed following the method of Koele and Hoogstraten (1988).

The authors are indebted to Dick de Gilder for his assistance. Address correspondence to Johan Hoogstraten, Methodology Unit, University of Amsterdam, Weesperplein 8, 1018 XA Amsterdam, The Netherlands.

Table 1
2 × 3 Analysis of Variance Results

Source	*SS*	*df*	*MS*	*F*	*p*
Between Subjects					
Groups	76.44	1	76.44	.08	.779
Subjects Within Groups	40,077.07	44	910.84		
Within Subjects					
Measurement	474.58	2	237.29	5.34	.006
Measurement × Groups	337.70	2	168.85	3.80	.026
Measurement × Subjects Within Groups	3,910.42	88	44.44		

DATA ANALYSIS

First, a 2 × 3 analysis of variance with repeated measures on one factor was carried out. Factors were groups (experimental, control) and measurement (pre, post, then). The results of this analysis are presented in Table 1. The measurement × groups interaction was significant, indicating that the so-called "null model" could be rejected; there evidently existed a response-shift bias and/or a treatment effect.

In the next step, the means in the experimental group were corrected for estimated systematic effects between pretest, posttest, and thentest in the control group. This was done by subtracting the grand mean of the control group from each of the three means in the control group. The resulting effects were subtracted from the corresponding means in the experimental group. In this way the experimental group means were corrected for systematic, nonexperimental effects due to testing, maturation, and so forth. Table 2 presents the uncorrected and the corrected means.

The corrected means of the experimental group formed the basis for the subsequent analysis. The models that were fitted on these means are the ones presented in Figure 1 and Table 2 in Koele and Hoogstraten (1988). Model 1 assumes that there is a treatment effect but no response-shift bias. Model 2 is the reverse of Model 1: a response-

Table 2
Cell Means Before and After Correction

	Pre	Post	Then
	Experimental Condition		
Uncorrected	62.40	61.76	55.20
Corrected	61.70	62.30	55.36
	Control Condition (Grand Mean = 58.30)		
Uncorrected	59.00	57.76	58.14
Effect	.70	−.54	−.16

shift bias, but no treatment effect, has occurred. Models 3 and 4 both represent the presence of a response-shift bias *and* a treatment effect. In Model 3 there is a relatively weak response-shift bias; in Model 4 there is a strong response-shift bias. The models were fitted on the corrected means by way of carrying out a one-way analysis of variance with repeated measures on the scores in the three experimental cells, followed by stepwise testing the fit of the contrast associated with each model. The coefficients (also presented in Table 2 of Koele and Hoogstraten, 1988) of the contrasts are:

Model 1: (−.5;1;−.5)

Model 2: (1;−.5;−.5)

Model 3: (0;1;−1)

Model 4: (.5;.5;−.1)

In Table 3 the results of the one-way analysis of variance on the corrected scores in the experimental group are summarized.

Models 1 and 2 were fitted by calculating the Pearson correlation between the contrast and the corrected cell means. Model 1 yielded a proportion of .32 of explained variance. Model 2 a proportion of .19. Evidently the models did not fit very well. The tests on the proportions of unexplained variance by these models (using the test statistic given in formula 2 of Koele and Hoogstraten, 1988) gave as respective results $F_1(1,48) = 11.40$ ($p = .001$) and $F_2(1,48) = 13.58$ ($p = .000$). Both models were rejected, and Models 3 and 4 were fitted. Model 3 explained 81% of the corrected between-cells variance; Model 4 explained 99%. Obviously, tests on the proportions of unexplained variance were unnecessary. Model 4 was finally selected as giving the best description of the experimental results. Thus, the conclusion must be that in this experiment a response-shift bias seriously impaired the evaluation of the treatment effect when pretest/posttest difference scores were used for estimating the treatment effect. Thentest/posttest difference scores, corrected for systematic nonexperimental effects as estimated in the control group, provided an unbiased estimate of the treatment effect. A 95% confidence interval for the mean of these corrected difference scores was (4.04; 9.84), or (.24; .58) on the 7-point scale of the test items. An evaluation of the relevance of this treatment effect stands apart from the model fitting procedure, and should be carried out in the usual way.

Table 3
One-Way Analysis of Variance Results

Source	*SS*	*df*	*MS*	*F*	*p*
Between Subjects	21,783.25	24	907.64		
Within Subjects					
Measurement	739.33	2	369.66	8.38	.000
Measurement × Subjects	2,117.31	48	44.11		

CONCLUSION

This application of the model fitting technique for analyzing retrospective pretest/posttest designs illustrates succinctly the straightforwardness of the procedure. Using the information supplied by the control group to its full advantage and separating response-shift bias from treatment effect, the proposed method enables the experimenter to estimate treatment effects that are free of some important sources of error variance.

REFERENCES

HOOGSTRATEN, J. (1982). The retrospective pretest in an educational training context. *Journal of Experimental Education*, **50**, 200-205.

KOELE, P., & HOOGSTRATEN, J. (1988). A method for analyzing retrospective pretest/posttest designs: I. Theory. *Bulletin of the Psychonomic Society*, **26**, 51-54.

(Manuscript received for publication June 8, 1987.)