Research Technique

The Slope Test: Applications in Formative Evaluation

Jon Baggaley
Aaron-Henry Brauer

Abstract: Logistical problems associated with formative evaluation reduce the reliability of pre-test/post-test comparisons as a basis for understanding educational effects, A paradigm shift Is recommended. focussing attention upon the adjusted post-test scores which are produced when the 'slope test' for homage neity of regression is conducted In a between-groups analysis of covariance.

PROBLEMS IN FORMATIVE EVALUATION

Formative evaluation is often a rough-and-ready process, in which the rigour associated with conventional research methods is unattainable. The essence offormative evaluation is to assess the impact of educational materials while there is still time for the production team to make modifications. The process must therefore be as speedy as possible. It may also have to be organized amid extreme pressures, especially in complex contexts such as TV production where a study can be required with little or no notice.

In this situation, a major problem is that of audience sampling. If the intended audience of the educational product is specific, the impact of the materials upon a representative audience sample can be assessed with relative ease. A simple post-test can establish the sample's overall reactions to the material; and the added use of apre-test can verify that specific knowledge was imparted by the educational materials (or impeded), or that attitudes/behaviour were changed (Campbell & Stanley, 1966). Pre-test scores in their own right can indicate that the audience was in possession ofcertain facts, attitudes or behavioural traits before the material was presented, and that aspects of the latter are therefore redundant.

As the target audience becomes more diverse, however – comprising both sexes, different age and educational levels, and wide-ranging attitudes or aptitudes – the effects of educational treatments upon it become harder to discern. At the pre-or post-test levels individually, even simple effects may go unnoticed, owing to sampling biases beyond the evaluator's control. The

CJEC. VOL. 18, NO. 1, PAGES 65 · 68, ISSN 0710 · 4340

problem is compounded when pre- and post-test scores are compared, since pre-to-post shifts may well be concealed by prior differences between separate audience subgroups.

In many evaluation studies, of course, it is possible to anticipate the critical audience variables on which treatment effects will depend. Audience subgroups can then be identified, and the pre- and post-test scores of each (e.g., men versus women) compared. Treatment effects on subjects with good reading ability, for example, may be compared with those observed on poor readers, and so on. The resulting evaluation scheme is the "pre-test/post-test multiple-group" design (PPMG), applied in aptitude-treatment interaction studies (Salomon, 1979).

APTITUDE VERSUS TREATMENT EFFECTS

In a PPMG context, the confounding effects of extraneous audience aptitudes can be identified via a statistical procedure known as the "slope test" (Cronbach & Snow, 1977). The test is particularly applicable in formative evaluation studies owing to the common logistical problems encountered in this area. It is necessary for the slope test manoeuvre that the pre- and post-test measures be precisely matched, with respect to content validity as well as to the individual subjects in the sample.

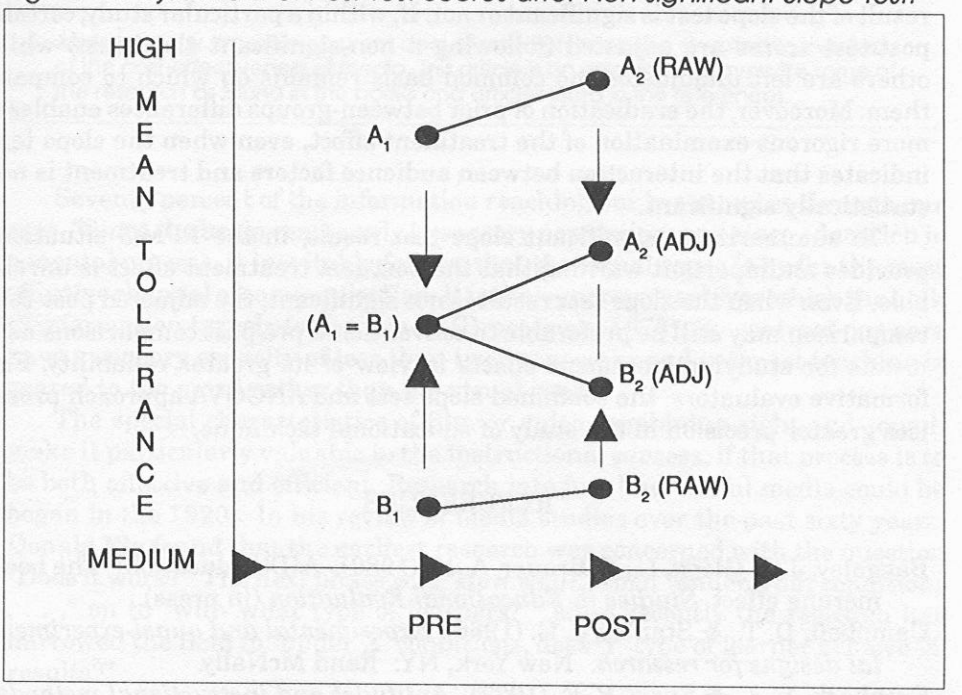
When, within a particular audience group, a significant treatment effect is present, the pre- and post-test scores of the group will be relatively uncorrelated, and the "slope" of the scores, plotted graphically, will be flat. If the treatment effect is dominated by audience aptitudes, however, the pre- and post-test scores of the group will be highly correlated, and the slope will be steep. If different treatment effects occur in separate audience subgroups, the slope test will indicate that the groups' pre/post slopes are significantly different, and that the treatment effect cannot be explored in terms of pre-to-post response shifts, owing to violation of the statistical assumption of homogeneity of regression.

In the PPMG context, the slope test thus provides the same safeguard against an unreliable treatment effect as the interaction term provides in one-way analysis of variance (Tabachnik & Fidell, 1983). If the result of the slope test is not significant, the evaluator may proceed to examine the educational treatment effect via an analysis of covariance (ANCOVA). With the pre-test scores as covariates, and audience differences as the independent variable, the sample's post-test scores are adjusted to take account of between-subgroups prior differences. Differences between the adjusted post-test scores of the subgroups can then be directlyattributedto the treatment. (N.B. The slope test and subsequent one-way ANCOVA are available within a single procedure when PPMG data are analyzed via the BMPDlV statistical package.

An Example

The utility of the slope test and covariance procedure was noted in a recent evaluation study of an educational film about AIDS (Baggaley, Glegg & Brauer, 1989). Figure 1 indicates the confounding observed in the study between the treatment effect (pre-test to post), and the attitude differences of separate audience subgroups (high-risk subjects vs. low-risk). The attitude measure in question was one of *tolerance* towards the rights of persons with AIDS in society. The slope test indicated that the pre/post effect of the film on this measure was nsignificantly different in the two AIDS-risk groups.

Figure 1. *Adjustment of Post-test Scores after Non-significant Slope Test.*



As the pre-test mean scores show, high-risk subjects (A) were markedly more tolerant than low-risk subjects (B). At the post-test level (raw means), this effect seems to have increased. However, the precise effect of the educational treatment on the post-test scores is not yet apparent. In order to isolate it, the between-groups pre-test difference was reduced via one-way ANCOVA to zero ($A_1 = B_1$), and the post-test mean scores (A_2 raw and B_2 raw) were adjusted accordingly. The treatment effect can now be inferred from the difference between the adjusted post-test means (A_2 adj – B_2 adj); in this instance the difference was found to be statistically significant. On this basis, Baggaley, Glegg & Brauer concluded that the educational film had polarized the attitudes of high and low-risk viewers with respect to social *tolerance*, and had thus done more educational damage than good.

APARADIGM SHIFT

The covariance analysis of pre/post-test data involves a style of questioning to which formative evaluators are unaccustomed. Instead of asking the usual question ("Are the pre and post-test responses different?"), the analysis asks *"Are the post-test responses of audience subgroups different, all pretest factors being equal? "The* "paradigm shift" from the first question to the second anticipates the statistical problems with which formative evaluation studies are plagued.

The identification of treatment effects via ANCOVA-adjusted post-test scores may be desirable in some formative evaluation situations whether the result of the slope test is significant or not. If, within a particular study, certain post-test scores are adjusted (following a non-significant slope test) while others are left unadjusted, no common basis remains on which to compare them. Moreover, the eradication of prior between-groups differences enables a more rigorous examination of the treatment effect, even when the slope test indicates that the interaction between audience factors and treatment is not statistically significant.

To summarize, a significant slope test result, in the PPMG situation, provides an important warning that the post-test treatment effect is unreliable. Even when the slope test result is not significant, the adjusted post-test comparison may still be preferable to conventional pre/post comparisons as a means for studying treatment effects in view of its greater reliability. For formative evaluators, the combined slope test and ANCOVA approach promises greater precision in the study of educational technique.

REFERENCES

Baggaley, J. P. Glegg, L., & Brauer, A. H. (1989). AIDS education: The boo-*merang* effect. *Studies in Educational Evaluation (in* press),

Campbell, D. T., & Stanley, J. *C.* (1966). *Experimental and quasi-experimen*tal *designs for research.* New York, NY: Rand McNally,

Cronbach, L. J., & *Snow,* R. E. (1977). *Aptitudes and instructional methods.* New York: Wiley.

Salomon, G. (1979). *Interaction of media, cognition, and learning. San* Francisco, CA: Jossey Bass.

Tabachnik, B. G., & Fidell, L. *S. (1983). Using multivariate statistics. New* York, NY: Harper & Row.

AUTHORS

Jon Baggaley is a Professor of Education, and Director of the M.A. Programme in Educational Technology at Concordia University, Montreal, Quebec.

Aaron-Henry Brauer is a Master's candidate in Educational Technology in the same Department.

ACKNOWLEDGMENT: The authors are indebted to Dr. Robert Bernard for statistical advice and insights supplied throughout the development of this paper.