dia exposure for the adoption of family planning practices (Winett, 1986). Cognition and performance *can* be influenced by the media (Bandura, 1986).

Thus, Costanzo et al. have provided a good overall model for inquiry and intervention in the general area of information and behavioral influence. However, their insistence that their current article and literature review "serves to underscore a well-established principle of persuasive communication" (p. 528), that is, the media are ineffective for behavior change, is problematic and may unfortunately deter more sophisticated research on the media and behavior change.

REFERENCES

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* New York: Prentice Hall.

Costanzo, M., Archer, D., Aronson, E., & Pettigrew, T. (1986). Energy conservation behavior: The difficult path from information to action. *American Psychologist, 41,* 521–528.

Manoff, R. K. (1985). *Social marketing: New imperative for public health.* New York: Praeger.

Rogers, E. M., & Kincaid, D. C. (1981). *Communication networks: Toward a new paradigm for research.* New York: Free Press.

Winett, R. A. (1986). *Information and behavior: Systems of influence.* Hillsdale, NJ: Erlbaum.

Winett, R. A., Leckliter, I. N., Chinn, D. C., Stahl, B., & Love, S. Q. (1985). Effects of television modeling on residential energy conservation. *Journal of Applied Behavior Analysis, 18,* 33–44.

Correspondence concerning this comment should be addressed to Richard Winett, Psychology Department, Virginia Polytechnic Institute, Blacksburg, VA 24061.

Evaluation Issues in a Quasi-Experiment on Teaching Thinking Skills

Irvin Sam Schonfeld
City College of New York

Herrnstein, Nickerson, de Sanchez, and Swets (November 1986) made a significant contribution to evaluating an educational intervention that was designed to advance the thinking skills of Venezuelan children attending seventh grade classes. As indicated by the mean scores on the psychoeducational tests administered to the students in the experimental and control groups before the intervention was implemented, the groups were not equivalent, with the youngsters assigned to the experimental group performing better than the control youngsters. This is unsurprising because the children were not randomly assigned to the groups; the evaluation constitutes a quasi-experiment (Cook & Campbell, 1979).

It was surprising to find that the principal method of evaluating the differences between the groups was the *t* test performed on pretest and posttest differences, or gain scores. Cook and Campbell (1979) pointed out a number of biases inherent in gain score analyses that need not be highlighted here. A major threat to the validity of the study, missed in a gain score analysis, is the possibility that an interaction between prior achievement or prior ability and group membership can account for the results.

In lieu of the *t* test, it would seem preferable to adopt multiple linear regression techniques as the chief analytic tool. In the approach I would envision, each posttest variable would be regressed on the corresponding pretest variable, group membership, which would be represented by a 0–1 dummy variable (Cohen & Cohen, 1983) and relevant control variables such as age and sex. A group-by-pretest product term would also be included in each regression equation to evaluate interaction effects.

The regression approach to assessing for an interaction is preferable to the approach taken by Herrnstein et al. (1986). They divided the groups into deciles based on pretest performance and examined the patterns of gains for experimental and controls subjects at each decile. Two problems emerge in connection with this approach. First, it is not clear if the decile rankings into which the experimental and control subjects were divided are based on the pooling of all the subjects' scores, or if each group was divided into deciles based on the subjects' within-group standings. If the latter condition is the case, the matched deciles are not equivalent, and the comparisons are biased in favor of the experimental group, which had the higher initial scores. If the former is the case, the comparisons are still likely to be biased. For example, given the possibility of "regression artifacts" (Campbell & Erlebacher, 1975), the *true* pretest scores of the experimental subjects with the lowest pretest performances are probably higher than the *true* pretest scores of the control subjects with the lowest pretest performances.

Second, Herrnstein et al. (1986) relied on the inspection of plots depicting percentage correct and percentage gain correct for experimental and control subjects at each decile to rule out the occurrence of an interaction. They did *not* perform any statistical tests to assess for interactions. They argued that

> not having adduced evidence for an interval scale of test scores, we shall not look for patterns in the functions shown in the bottom panels of Figures 2 through 5, for example, in their slopes. The general appearance is of slight differences in gain across ability levels for a given test and only unsystematic differences from test to test. (p. 1285, fn.)

The disclaimer about "interval" data is weak because their reliance on gain scores suggests that they did indeed treat test scores as interval data. At least a regression approach to the data would allow for systematic tests of group-by-prior-achievement and group-by-prior-knowledge interaction hypotheses.

The multiple regression approach has an additional advantage. If no interaction is found, the unstandardized regression, or "b" weight, for group membership takes on a convenient meaning, namely, average posttest advantage associated with membership in the experimental group, controlling for initial differences in achievement. Moreover the estimates of the "b" weights are likely to be minimally biased (Kenny, 1979) because of the high reliability of each of the achievement measures employed, the Otis-Lennon School Ability Test and the General Achievement Test. Although reliability data on the tests developed expressly for the quasi-experiment, the Target Abilities Tests, are not presented, the sheer numbers of items involved suggest that the tests are reliable.

Herrnstein et al. (1986) presented the intertest correlations, which show that gain scores correlate with each other less strongly than either the pretests or the posttests. They argued that "the much lower correlations for gains confirm the earlier conclusion that ability levels were only weakly associated with changes in scores during the year" (p. 1286–1287). However, they ignored an important rival, and more plausible, explanation of the lower correlations between gain scores, namely, that the reliability of the gain scores is lower than that of the pretest or posttest scores (Cohen & Cohen, 1983).

REFERENCES

Campbell, D. T., & Erlebacher, A. E. (1975). How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In E. Struening & M. Guttentage (Eds.), *Handbook of evaluation research* (pp. 597–617). Beverly Hills, CA: Sage.

Cohen, J., & Cohen, P. (1983). *Applied multiple*

regression/correlation analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin.

Herrnstein, R. J., Nickerson, R. S., de Sanchez, M., & Swets, J. A. (1986). Teaching thinking skills. *American Psychologist, 41*, 1279–1289.

Kenny, D. A. (1979). *Correlation and causality*. New York: Wiley.

A Code of Practice for Refereeing Journal Articles

James Hartley
University of Keele

Many authors have commented on aspects of journal refereeing (e.g., Ceci & Peters, 1984; Harnad, 1982; Standing & McKelvie, 1986; Surwillo, 1986). Gordon (1980) and Franzini (1987) made the important point that, even if journals use anonymous refereeing procedures, the editors choose the referees. Presumably the argument is that editors can bias the decision whether to accept or reject a paper by their choice of referees, and that this can happen whether or not anonymous refereeing is used.

In order to incorporate the many suggestions made for improving publication procedures, it seems that one needs a general code of practice to which all journal editors might aspire. I tentatively suggest the following: (a) Editors use anonymous refereeing; (b) referees rate articles for their significance, value, and so on on standard rating scales, and then add their comments; (c) referees sign their reports; (d) editors send the full referees' reports to authors; (e) editors send the full referees' reports to each of the referees involved; and (f) each journal sets up an independent group to discuss appeals if arguments between editors and authors cannot be resolved.

Such an agreed-upon code of practice would not eliminate all the problems associated with journal refereeing, but it would solve some of the difficulties. Anonymous refereeing, although not perfect, is more evenhanded. Standard rating scales make for easier comparisons between different judges (and different articles). Referees' signing of reports should lead to better quality reviewing. Sending signed referees' reports to the authors should overcome Franzini's criticisms to some extent. Seeing other referees' reports is informative for lone referees. Finally, allowing for the possibility of an appeal would make for a fairer system.

No doubt the editors of many APA journals already practice some parts of this suggested code. However, would things not be better if all of it were followed?

REFERENCES

Ceci, S. J., & Peters, D. (1984). How blind is blind review? *American Psychologist, 39*, 1491–1494.

Franzini, L. R. (1987). Editors are not blind. *American Psychologist, 42*, 104.

Gordon, M. D. (1980). The role of referees in scientific communication. In J. Hartley (Ed.), *The psychology of written communication: Selected readings* (pp. 263–275). New York: Nichols.

Harnad, S. (1982). *Peer commentary on peer review: A case study in scientific quality control*. New York: Cambridge University Press. (Reprint of *The Brain & Behavioral Sciences, 5* [2])

Standing, L., & McKelvie, S. (1986). Psychology journals: A case for treatment. *Bulletin of the British Psychological Society, 39*, 445–450.

Surwillo, W. W. (1986). Anonymous reviewing and the peer review process. *American Psychologist, 41*, 218.

Correspondence concerning this comment should be addressed to J. Hartley, Department of Psychology, University of Keele, Keele, Staffordshire, England ST5 5BG.