Reply to Morley

DANIEL J. O'KEEFE
University of Illinois
SALLY JACKSON
SCOTT JACOBS
University of Oklahoma

Morley's reply offers three main claims: first, that multiple-message study designs "will not necessarily eliminate threats to external validity"; second, that there is "no value in treating systematically constructed message samples as random"; and, third, that generalization should be based not on statistical inference but "on logical grounds."

The use of multimessage designs. Morley is correct that the use of multiple messages will not eliminate all threats to validity. In particular, a manipulation that *builds in* a confounding between the variable of interest and some other variable (e.g., building length into sidedness by uniformly *adding* arguments to create two-sided messages) will not be corrected through application to multiple messages. This situation can be improved only through careful reflection on the match between the variable as conceptualized and the manipulation used to operationalize it.

However, the problem Morley describes is not the problem that Jackson and Jacobs (1983) offered to remedy. Jackson and Jacobs were primarily concerned not with homogeneous and systematic confounds associated with ill-conceived manipulation procedures but with heterogeneous and idiosyncratic variations introduced by the specific concrete messages chosen to instantiate the features of interest. This problem *can* be mitigated or solved outright through the use of multiple messages.

The key question is not whether multiple-message designs solve all of our design problems, but how multiple-message designs compare with the alternative—single-message designs. If a method of manipulation

consistently confounds one variable with another, neither replication nor statistical maneuvering will help. Nevertheless, there does not seem to be much doubt about the palpable inadequacy of single-message designs as bases for the generalizations they are usually intended to support, nor about the superiority of multiple-message designs. Whatever problems are left unresolved by the sheer act of replicating messages, multiple-message designs are much preferable to single-message designs.

The analysis of multiple-message designs. Morley disclaims an interest in defending single-message designs, but offers no proposal for analyzing multiple-message designs. Our own view is surely clear enough: In most cases, messages should be treated as replications, that is, as levels of a random factor. Morley says that he finds "no value in treating systematically constructed message samples as random," but does not recommend any alternative analytic strategy.

The reasserted objections to treating messages as levels of a random factor appear to be (a) the nonrandom construction of messages and (b) considerations of practicality and power.

Concerning the nonrandom construction of messages: Morley's reasserted objection emphasizes the acknowledged impossibility of guaranteeing the removal of all possible confoundings between the variable of interest and other factors. But this impossibility does not underwrite treating messages as a fixed factor. That such confoundings may occur does not remove the possibility that the particular messages studied contribute to the variability of treatment effects, and hence does not diminish the desirability of treating messages as replications.

Morley argues that since the messages (which obviously aren't randomly sampled) cannot be randomly assigned to treatments, treating them as a random factor is inappropriate. But communication researchers customarily treat subjects as random in at least two sorts of cases in which neither random sampling nor random assignment occurs: first, when subjects are classified on individual-difference variables such as gender, and, second, when each subject gets every level of the treatment (i.e., repeated measures designs). Obviously, researchers don't require random assignment (or random sampling) of subjects as prerequisite to treating subjects as a random factor, and hence it is unreasonable (and arbitrary) to insist that random sampling or random assignment are required for treating messages as a random factor.

Morley's premise is in fact false: It *is* possible to assign messages randomly to treatments (though we emphasize that this is not necessary to underwrite treatment of messages as a random factor). Instead of crossing messages with treatments, an experimenter could randomly assign base messages to levels of the treatment before applying the relevant treatment modification—with the effect that messages (like subjects) would be nested within treatments. We are *not* recommending such designs—quite obviously, they do not improve either the statistical or the logical grounds for generalization—but merely pointing out that Morley's stated reasons for treating messages differently from subjects can be dismissed.

Concerning practicality and power: Researchers recognize that small subject samples make for poor power, and that many subjects may be needed to have good power—but no one thinks these considerations justify treating subjects as a fixed factor. We may now recognize that small numbers of messages make for low power, and that many messages may be needed to have good power—but these considerations are not good reasons for treating messages as a fixed factor. If we acknowledge that comparisons based on many messages are preferable to comparisons based on a few messages (as surely we should), then we should not be surprised that obtaining these more desirable comparisons will incur costs.

High power might be obtained by using single-message designs, but of course we are talking then about the power to detect treatment effects for one particular message—disregarding entirely both Type I and Type II errors concerning treatment effects for a class of messages. Nor does this apparent gain in power compensate for the patent inadequacy of such designs as sources of evidence for the generalizations they are usually intended to support. Morley's discussion of "the power of a meta-analysis" is quite misleading, for in fact his discussion is limited to combined probability procedures for meta-analysis (of the sort discussed by Strube and Miller, 1986), and does not concern meta-analytic techniques based on effect sizes (the preferable meta-analytic approach; see Becker, 1987). It is not true that the power (or, more generally, the sensitivity) of all meta-analytic procedures "stems almost exclusively from the total number of subjects contained in the combined studies and not from the number of studies combined" (Morley, this issue), as examination of various meta-analytic procedures will reveal (Osburn, Callender, Greener, & Ashworth, 1983, Table 6; Hedges & Olkin, 1985, p. 199).

Generalization on logical grounds. According to Morley, the role of meta-analysis is not to warrant generalizations but to indicate that "a body of research has found a given effect at some level of probability" (see Note 1 to our earlier reply). However, he believes one can generalize "on logical grounds," along the lines of "twenty studies conducted by thirty researchers using sixty different messages collectively indicate treatment A to be more persuasive than treatment B." But Morley leaves unanswered many important questions presupposed by such a "generalization on logical grounds": (1) For those using multiple messages, how is the comparison of treatments made? (2) How many of the messages within a given study have to favor treatment A before the study is counted as favoring treatment A? (3) How many studies have to favor treatment A before the studies "collectively" favor treatment A? And (4) how much do the two treatments have to differ before we are willing to interpret the difference at all? If these questions are addressed with statistical evidence, Morley is vulnerable to precisely the arguments he levels against Jackson and Jacobs. And if statistical evidence is ruled out, it is not clear what alternatives are available.

Sooner or later, the issue of how to analyze multiple-message data has to be faced squarely. Employing single-message designs pending meta-analyses will only defer, not avoid, the question of the treatment of messages as random (because the question reappears in the context of meta-analysis). Our view is that because message replications are a potential source of error in the evaluation of categorical effects, those replications should be treated as a random factor.

REFERENCES

Becker, B. J. (1987). Applying tests of combined significance in meta-analysis. *Psychological Bulletin, 102,* 164-171.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* New York: Academic Press.

Jackson, S., & Jacobs, S. (1983). Generalizing about messages: Suggestions for design and analysis of experiments. *Human Communication Research, 9,* 169-191.

Osburn, H. G., Callender, J. C., Greener, J. M., & Ashworth, S. (1983). Statistical power of tests of the situational specificity hypothesis in validity generalization studies: A cautionary note. *Journal of Applied Psychology, 68,* 115-122.

Strube, M. J., & Miller, R. H. (1986). Comparison of power rates for combined probability procedures: A simulation study. *Psychological Bulletin, 99,* 407-415.