The Search for Reliable Generalizations About Messages
A Comparison of Research Strategies

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

Two ways of conducting the search for generalizations about messages are considered: Morley's (this volume) proposal that single-message research designs be used, with subsequent meta-analytic summaries, and Jackson and Jacobs's (1983) proposal that multiple-message designs be used, with messages treated as a random factor in the statistical analysis. Jackson and Jacobs's approach is shown to provide a more dependable, efficient, and practical means for gathering the requisite evidence for dependable generalizations. The charge that multiple-message designs suffer from irreparable problems of experimenter bias is refuted. The treatment of messages as a random factor is defended as statistically appropriate and as clearly preferable to the statistical alternatives.

This essay considers two ways of conducting the search for generalizations about messages, one advanced in preliminary form by Jackson and Jacobs (1983), and the other advanced along a broad front but championed most recently by Morley (this volume). We will describe and compare the two approaches, then respond to specific objections raised by Morley against the Jackson and Jacobs proposal.

Sally Jackson (Ph.D., University of Illinois, 1980) is an associate professor of communication at the University of Oklahoma. *Daniel J. O'Keefe* (Ph.D., University of Illinois, 1976) is an associate professor of speech communication at the University of Illinois. *Scott Jacobs* (Ph.D., University of Illinois, 1982) is an associate professor of communication at the University of Oklahoma. The authors wish to thank Barbara J. O'Keefe for many helpful suggestions.

Human Communication Research, Vol. 15 No. 1, Fall 1988 127-142

MESSAGE GENERALIZATION: TWO RESEARCH PROPOSALS

Despite some appearances to the contrary, Morley shares with Jackson and Jacobs an underlying interest: The establishment of dependable generalizations about messages.[1] Generalizing about messages involves finding differences between message classes or relationships among message characteristics (e.g., "Two-sided arguments are more persuasive than one-sided arguments").

Dependable message generalizations are not easily achieved, however. The typical research strategy has been to draw categorical conclusions from studies in which the categories of interest are each represented by a single message (Jackson & Jacobs, 1983). But such studies rarely if ever justify the conclusions drawn from them: They confound the individual message case with the general message category, and such case-category confounding threatens both internal and external validity.

The internal validity problem arises because any two individual messages may differ in an indefinitely large number of ways. As a result, an observed difference (e.g., in persuasiveness) between two messages cannot reasonably be attributed to any one categorical difference. Recognizing this, experimenters commonly attempt to "control" or "hold constant" everything except the message variable of interest, by manipulating this variable within the context of a single base message. Although this controls the grosser aspects of the problem, it is not a solution; it merely shifts the locus of the problem, since whatever message elements instantiate the manipulation inevitably represent not one clear contrast but a bundle of contrasting characteristics.

For example, if two alternative experimental paragraphs are inserted into a common base message so as to manipulate a variable such as argument sidedness, the problem occurs within the experimental paragraphs: The two paragraphs may differ not only on sidedness (the variable of interest), but also on an indefinitely large number of other variables. The paragraphs may differ in logical soundness, or language intensity, or quality of evidence, or length, or imagery, or repetitiveness, or any number of other attributes, and hence any difference in persuasiveness might be explained just as well by one of the unintended manipulations as by the intended manipulation. Even if the manipulated elements are not paragraphs, but mere sentences or individual words (as in the manipulation of language intensity), these problems remain (as noted by Ellis, 1982).

But even if the experimental treatment could be applied in such a way as to isolate the intended manipulation, such an experiment would offer little strong basis for categorical claims, because external validity problems would remain. The external validity problems derive from the possibility that the manipulation's effect will vary across messages. Unless we assume that the effect of the message treatment is uniform across messages (e.g., that two-sidedness is uniformly advantageous or disadvantageous for all the messages in which it might be used), a comparison made for one message cannot be considered an adequate basis for guessing the effect of the treatment on other messages.

To address these barriers to generalization, Jackson and Jacobs (1983) proposed that all experiments on message variables incorporate multiple comparisons. Any message category or message treatment defined by the design should have multiple-message replications, to prevent or mitigate confounding of the message variable with other unsuspected variations in the messages. These replications may be nested within categories or crossed with treatments—whatever the research problem requires. And further, these replications should be recognized as a source of error in the evaluation of categorical effects—that is, message replications should be treated as a random factor.

To build a case for a generalization about a variable such as messagesidedness, the basic plan would be to gather a large number of base messages (with the number—henceforth, M—as open to variation as is the number of human subjects, N), to apply the experimental manipulation of sidedness to each of the base messages to produce a one-sided and a two-sided version of each, and to present each of these manipulated messages to an audience of randomly assigned human respondents. Messages would be considered a random effect, so the effects of sidedness would be tested against the message x sidedness interaction.

A significant difference between the two-sided versions and the one-sided versions would be interpreted like any other significant treatment difference. Such a result will tell us that the overall treatment difference is large enough relative to the variability of treatment differences from replication to replication to convince us that one of the two treatments is the better.

Morley proposes an alternative research strategy based on meta-analysis of single-message studies. Under his proposal, the basic pattern for an individual experiment would be exactly as in the existing literature: Each study would apply the sidedness treatment to a singlebase message (i.e., M = 1). However, interpretations of the results of these studies would be very limited, and general claims of the sort desired ("Two-sided argumentation is more persuasive than one-sided argumentation") would be advanced only after the accumulation of numerous independent experiments. A meta-analysis of 10 M = 1 experiments with N subjects each would offer about the same level of interpretability as a single M = 10 experiment with 10N subjects.

A COMPARISON OF THE TWO APPROACHES

Notice that both proposals can enjoy the undisputed benefits of meta-analytic techniques. After all, meta-analysis can be done on bodies of multiple-message studies as readily as on bodies of single-message studies. Morley's proposal enjoys no unique advantages (concerning, e.g., bias detection) because of meta-analysis. The issue in dispute thus comes down to how primary research on message variables should be conducted.

Addressing the Barriers to Generalization

Initially, we think Morley's approach inferior because it misapprehends the problems connected with message generalization. The problem of message generalization, for Morley, concerns only the *scope* of the conclusions drawn (the external validity problem): His essay suggests that the difference between two-sidedness and one-sidedness can be assessed in a single-message study, but cannot be assumed to hold for messages other than the one observed.

But this overlooks the internal validity problems described above, for it fails to appreciate that the absence of multiple messages creates problems for the *formulation* of categorical claims. For example, even if we are (apparently) very conservative in our interpretation of a single-message study of argument sidedness, and conclude only that two-sidedness is advantageous for the particular topic or message observed, we have still made an unwarranted categorical claim, because sidedness may not be the only important categorical difference between the experimental messages. Since we have no effective way to manipulate one message variable without affecting others, only multiple comparisons offer any strong basis for any particular categorical claim.

Dependability and Efficiency

Additionally, Jackson and Jacobs's proposal is a more efficient and dependable way of gathering the desired evidence for message generalizations. Morley's proposal depends upon the availability of a large number of separate studies bearing on the same research question. Like primary analysis, meta-analysis suffers from low power and poor reliability when the number of independent studies is few. This dependence on a large number of separately conducted studies may be a crippling disadvantage, since no individual study can be planned or published with any assurance that it will lead to a large number of replications. But without replications, Morley's proposal collapses, since individual studies are uninterpretable. Moreover, even assuming the availability of a large number of studies for meta-analysis, Morley's proposal encounters difficulties familiar to any who have conducted meta-analysis: inadequacies in statistical reporting; inadequacies in design information; nonindependence among the studies; the suspicion of publication bias; and inaccessibility of even the studies known to exist.

Unlike Morley's proposal, the Jackson and Jacobs proposal provides that each study will contain the raw materials of generalization: multiple examples of the categories of interest. An individual study designed to Jackson and Jacobs's specification can support a reasonable argument for a general categorical claim, while an individual study designed to Morley's specifications can never do that.

Practicality

Morley argues that incorporating multiple messages imposes an unreasonable burden on the experimenter or on participants or both. Apart from the fact that general decisions about crucial design issues ought not be made on the basis of convenience to the researcher, Morley's argument is based on misrepresentation of what Jackson and Jacobs recommended. Jackson and Jacobs do not demand that every subject respond to every message, nor do they suggest all that research questions will allow messages to be confounded with subjects. Jackson and Jacobs suggest only that messages be replicated; many implementations are possible.

Many recent studies have incorporated multiple messages (Bradac & Mulac, 1984; Cody et al., 1986; Doelger, Hewes, & Graham, 1986; Hample & Dallinger, 1987; Hewes, Graham, Doelger, & Pavitt, 1985;

Jackson & Backus, 1982; Jackson, Jacobs, Burrell, & Allen, 1986; Jackson, Jacobs, & Rossi, 1987; Mulac, Bradac, & Mann, 1985; Mulac, Lundell, & Bradac, 1986; O'Keefe & McCornack, 1987; Planalp, 1985; Planalp, Graham, & Paulson, 1987; Tracy, 1982, 1983, 1984; Tracy, Craig, Smith, & Spisak, 1984). Not all of these are fully realized examples of the Jackson and Jacobs proposal, but all represent practical improvements over single-message designs.

TWO OBJECTIONS TO THE JACKSON AND JACOBS PROPOSAL

Morley offers two substantive objections to the Jackson and Jacobs proposal, corresponding to the two main components of that proposal: the design suggestion that experiments should incorporate multiple messages and the statistical suggestion that messages be treated as random effects.

Should Studies Contain Multiple Messages?

Even though Morley concedes that generalization from a single-message study is unwarranted, he defends single-message studies over multiple-message studies. Morley argues that any experimental message must reflect the "message-generating biases" of the experimenter. From this he concludes that multiple messages are nonindependent and compound whatever biases are present in single-message studies. Potential problems of bias and nonindependence are important ones, of course, and deserve close attention.

However, Morley's remedy (meta-analysis of single-message studies) is no help. Bias and nonindependence are equal (or worse) problems for meta-analysis (cf. Rosenthal, 1984, pp. 125-127). Examination of major lines of research on message variables will reveal not only the sort of nonindependence that worries Morley (multiple studies done by a single experimenter) but even worse problems of bias and nonindependence (e.g., one experimental message appearing in study after study, or later studies modeling their messages after those in earlier studies). Morley's argument could easily be turned on his own proposal: Jackson and Jacobs's proposal controls the sorts of biases that otherwise might come about when results for one message are allowed to affect decisions about how to construct the next message. Thus even if the problem Morley notices were insoluble, it would support an argument against meta-analysis of single-message studies rather than an argument against multiple-message designs.

Happily, the problem is not insoluble, but can be tackled in a constructive way within the framework of multiple-message studies. One possible strategy (suggested by Jackson & Jacobs, 1983) is that experimental messages be sought among naturally occurring messages. Such a strategy is aimed at avoiding a sample of messages all written to fit a single pattern. For instance, a researcher interested in one-sided and two-sided argumentation might collect arguments from editorials, Letters to the Editor columns, direct mail ads, or student speeches, filling out the cells of the design by constructing two-sided versions of the one-sided originals and one-sided versions of the two-sided originals.

Stronger safeguards are easily devised. Researchers can arrange matters so that they have no active role in the production of base messages or in the creation of experimental manipulations. For example, messages can be solicited from colleagues or (if bias from professional expertise is a concern) from students or naive sources (for examples, see Mulac et al., 1986; O'Keefe & McCornack, 1987). Of course, every research situation is different, and no strategy provides a universal solution. But the special needs of individual cases will surely suggest other solutions.

Should Messages Be Treated as Levels of a Random Factor?

Morley argues that treating messages as a random factor is inappropriate, since it assumes that messages have been sampled at random from the population to which we wish to generalize. Asserting that "methodologists universally recognize that when researchers nonrandomly select the levels of a variable the variable is fixed," he develops an argument for the impossibility of generalizing about message populations. Morley is mistaken about what methodologists recommend for such situations, and he is wrong to assume that statistical inference depends on random sampling from a predefined population.

Must replications be sampled at random to be treated as a random factor? The problem of what to do with nonrandom replications is a recurring one. In communication research and many other fields, the problem arises with respect to human subjects, and most experimenters seem very little impressed with the suggestion that subjects really should be sampled at random.

The issue of what to do with replication levels chosen nonrandomly was addressed in detail 30 years ago by Cornfield and Tukey (1956), who argued that even if replication levels cannot be chosen randomly, they should be taken into account as a source of sampling error in estimating the differences among the (fixed) treatment levels. The same issue has enjoyed a recent revival among meta-analysts, including Glass, McGaw, and Smith (1981) and Hedges and Olkin (1985). Anticipating objections to the computation of inferential statistics in meta-analysis, Glass et al. defend the treatment of individual studies as replication factors even though it is known that they do not represent a random sample from a population of possible studies. Hedges and Olkin (1985, p. 190) offer "random effects models for effect sizes" that seem hand-tailored to the meta-analysis of message research, based on the supposition that each study "is a sample realization from a universe of related treatments" and that "'replication' of a treatment across [studies] may yield many different treatments, each sampled from some universe of possible treatments." That the "sample realizations" of the treatment variable cannot be randomly sampled from a population of possible treatments does not bother them at all.

Among researchers concerned specifically with messages, Clark's (1973) discussion of this issue is already familiar. But Morley represents Clark (1973, 1976) as arguing only for the treatment of *words* as random (defensible, in Morley's view, since words can be sampled from a dictionary), when in fact Clark (1973, p. 348) argues clearly and consistently that language materials of all sorts be so treated:

> When should the investigator treat language as a random effect? The answer is, whenever the language stimuli used do not deplete the population from which they were drawn. Note that the answer is *not*, whatever the language stimuli used were chosen *at random* from this population. The latter requirement is, in a sense, secondary to whether or not language should be treated as a random effect.

Clark's subsequent remarks make perfectly plain that his argument does not hinge on the possibility of sampling materials from a known population (see Clark, 1973, p. 352; 1976, p. 257, n. 2, and p. 260).

Even Keppel, who at first articulated a position very parallel to Morley's (Keppel, 1973, 1976), now acknowledges that the decision to treat the levels of a variable as fixed or random should be made not on the grounds of how the levels are chosen but on the grounds of how the levels function within the analysis (compare the 1982 edition of Keppel's textbook, especially pp. 519-520 and 533-537, with the 1973 edition cited by Morley). Wickens and Keppel (1983, p. 304) are also quite clear on this subject:

> As long as a nonsystematic sampling procedure is used, in which accidental confounding is a potential possibility, the random-effects model, which attempts to assess the magnitude of this confounding, is a better representation than the fixed-effects model, which ignores it.

The position that treating such nonrandom replications as levels of a random factor is preferable to treating them as fixed is shared not only by Cornfield and Tukey, Hedges and Olkin, Clark, and Wickens and Keppel, but also by Coleman (1964), Fontenelle, Phillips, and Lane (1985), Richter and Seay (1987), and Santa, Miller, and Shaw (1979). Contrary to what Morley believes, the appropriateness of treating messages as random (even when not randomly selected) has broad, authoritative support—and for good reason. The key question is whether estimates of treatment effects are assumed to be free of error due to the message sample: If we admit the possibility that the specific messages studied contribute to the variability of treatment effects, messages should not be treated as fixed.

This argument has implications beyond the specific issue of whether *messages* may properly be treated as levels of a random factor, since communication researchers commonly attempt statistical inference from samples of *people* selected nonrandomly from the population of interest. Morley suggests that these inferences too are inappropriate.[2] He argues as though the role of statistical inference in primary research can be only to generalize from a sample of observed cases to a predefined population of unobserved cases, even though he asserts its role within meta-analysis to be something else entirely, something free of assumptions about sampling and of implications for populations. This is not only inconsistent, but it is also faulty doctrine. Statistical tests are used in experiments to assess the strength of the evidence for a categorical claim, not to generalize a description of a sample to any sort of specifiable population, a distinction that is readily (if tacitly) understood by most experimenters (see Edgington, 1966). The mathematical basis for statistical testing is not irrevocably bound to notions of random sampling (see, for example, Finch, 1976). On the contrary, such mathematical issues as estimating the variance of a mean can be settled quite abstractly, with results adaptable to many configurations of assumptions about the observations to be analyzed.

Are alternative statistical choices suitable for multiple-message designs? One might accept Jackson and Jacobs's design suggestion that multiple messages be incorporated, but opt for a statistical treatment in which messages are not considered a random factor. At least three alternatives can be imagined, all better than a single-message design but all worse than a multiple-message design treating messages as random.

First, messages can be treated as an explicit but fixed factor (see, e.g., Housel, 1987; Rogers & Mewborn, 1976). Motivating this strategy is the hope of showing that messages do not interact with treatments, for then the treatment effect might be assumed to be constant across messages. Although this is an appealing line of argument, it has inherent weaknesses. Most obviously, the number of different messages examined has no effect on the power of any test of treatment effects, a consequence that clashes with our general sense that a comparison based on few messages is not as good as a comparison based on many. Circumstantially, when message *do* interact with treatments, this approach offers few resources for the analysis of treatment effects, short of "simple main effects" computed separately for each individual message (a predicament shared by both studies cited). Still, this approach ought not to be ruled out ipso facto, since lack of message x treatment interaction offers some evidence of generality, limited primarily by the number of message replications and by statistical power.[3]

A second alternative is to ignore messages entirely in the statistical analysis, so that, say, a treatment x message design collapses into a one-factor analysis. Such an approach offers no statistical information on generality, and is preferable to an unreplicated design only in that it mitigates to some extent the impact of variability among individual cases on the apparent differences among categories of messages, that is, it makes the treatment difference more reliable. In addition to all of the problems afflicting fixed-effects treatment of messages, this second alternative amalgamates variance due to messages, subjects, and message x treatment interactions, an outcome that is unacceptable unless all message effects are nil.

Third, following Morley's general line of reasoning, we might treat each message replication as a separate study and meta-analyze the study-specific results. Whether this differs in principle from the Jackson and Jacobs proposal is not clear, since any test of the significance of the average effect size treats the replications as random. (Note that meta-analysis does not avoid any assumptions about randomness, but slips them in under a broad umbrella.) The comparison of this option with the Jackson and Jacobs proposal comes down to details of the computational procedures. We might note, for example, that meta-analyzing such an experiment precludes pooling of within-groups variance and also precludes direct testing of the message x treatment interaction.[4] These costs are offset by no known benefits.

Are random-effect analyses statistically defective? Morley alludes to "statistical problems" associated with the treatment of messages as random effects. Chief among alleged problems are mathematical problems with the quasi-F ratio and the low power of multiple-message designs.

To be sure, the quasi-F is not uncontroversial, and individual researchers will want to examine the relevant literature. But one should not overlook the point made by Richter and Seay (1987, p. 478) that the alleged conservatism of the quasi-F test is surely preferable to the alternative, which in their view amounts to a decision to omit "any test of the hypothesis that there are no differences among conditions significantly greater than would be expected from random variation in subjects and words alone."

In any case, investigators leery of the quasi-F can easily design suitable experiments that avoid the need for quasi-F ratios. Morley falsely implies that quasi-F tests can be avoided only through nesting of messages within subjects or through confounding of subjects with messages. In fact, nesting subjects within messages (a standard independent groups design) will do just as well, and this is precisely the arrangement most suitable for studies requiring lengthy messages such as speeches or essays. Avoiding quasi-F is easy: Just nest subjects within messages, or nest messages within subjects, or confound messages and subjects—whatever the situation requires.

Concerning power: When messages are treated as random, the power of any test of treatment effects does indeed depend upon the number of messages included. If few messages are used, power will be low. But this is an argument for the use of many messages, not an argument for treating messages as fixed.

Morley implies that Wickens and Keppel (1983) see fixed-effects analysis as a reasonable alternative when the number of replications is so low as to threaten power. They do not. Noting that power can be increased under certain conditions by either treating materials (messages, words) as fixed (yielding the ratio labeled F_1 in both Clark, 1973, and in Wickens & Keppel, 1983) or treating subjects as fixed (the ratio labeled F_2), Wickens and Keppel remark:

> Unfortunately, this increase is of no value to an experimenter, for it is largely attributable to the accidental confounding of the treatment effects with the unconsidered error source, materials for F_1 and subjects for F_2. (pp. 303-304)

Similarly, Morley represents Wickens and Keppel as supportive of the view that low power results from treating messages as random, a representation difficult to reconcile with their article:

> It is important to design the experiment so that a sufficient number of levels is allocated to every source of random variation. It is certainly wasteful to allocate valuable resources to increasing the number of subjects when what is needed to improve power is a more generous sampling of materials. A failure to appreciate this point has led some researchers . . . to overemphasize the power deficiencies of the F' statistic. (p. 304)

The general point is this: Messages affect power in the same way as other replication factors, including subjects. Small numbers yield low power. But that does not justify treating messages as a fixed effect and it certainly does not justify studying messages one at a time. (Imagine a parallel argument that we should treat subjects as a fixed effect or study them one at a time!) If the appropriate statistical treatment of messages—as a random effect—makes for low power in a given design, the experimenter should boost power in legitimate ways (e.g., by adding message replications), not by treating messages as fixed effects.

Summary. Given multiple-message replications, the treatment of messages as a random effect is justified. The fact that messages cannot be randomly sampled is no objection to treating messages as a random factor or conducting statistical inference with messages considered as random; the statistical alternatives to treating messages as random are unattractive; and the putative statistical problems of random treatment of messages are entirely avoidable.

CONCLUSION

Morley's proposal—that experiments be built on single messages and that generalization be reserved for meta-analysis—is not reasonable. The argument that messages cannot legitimately be treated as random lacks sound foundations, and the charge that multiple-message designs necessarily amplify experimenter bias is false. Experiments on message variables can and should incorporate multiple messages, and, in most cases, these messages should be treated as replications, that is, as levels of a random factor.

But as this colloquy presumably makes clear, generalizing about messages raises many important issues that are not easily resolved. Some implications of the Jackson and Jacobs proposal remain unexplored, and undoubtedly some practical problems will arise. The development of strategies for generalization about messages must unfold through detailed and sustained reflection on individual empirical claims and the arguments offered in their support.

NOTES

1. At least sometimes Morley gives the impression that he has no interest in generalization. But if that's the case, it is not clear what to make of Morley's concern for experimenter bias in message construction. Experimenter bias is not a "problem" if one's interest is confined to the messages under study. Similarly, Morley's seeming reluctance to acknowledge openly his concern with generalization is reflected in confusing statements such as: "Meta-analysis simply permits the analyst to claim at some level of probability that a result was obtained in a sample or population of studies." But if one's interest is simply in seeing what the mean effect size *is* in the studies reviewed, one can simply compute that mean without performing any significance test and know that the probability is 1.00 that that is the mean for those studies.

2. Morley claims that "human samples can at least approach quasi-randomness," but the sense of this claim is unclear. A public opinion survey of a specified population might "approach quasi-randomness," but what of an experiment on the effect of argument-sidedness? In this case, the experimental hypothesis concerns the effects of a message strategy on *anyone anywhere at any time*. The relevant "population" includes persons not yet born and those already dead. In fact, it is not really a concrete collection of individuals at all, any more than a message category is a concrete collection of messages. The point is that random sampling of people for experimental research is not possible in any deep sense.

3. Some have suggested that this strategy provides a valid test of categorical effects that can then be generalized "nonstatistically." Unfortunately, this is not the case. Consider the application of such reasoning to an experiment comparing two treatments of M base messages. One might suppose that a significant difference between treatments would be sufficient to establish the categorical claim for the message sample observed, and that generalization to a broader class of messages might be defended on nonstatistical grounds. But treating messages as a fixed effect prevents evaluation of the categorical claim, even for the cases observed, producing nothing to generalize. Consider: A significant advantage for one treatment level might appear *by virtue of a single deviant message within the set of 2M*. One might well regard such a pattern as

evidence that something about the deviant message was peculiar (for example, it operationalized the message variable poorly), and could remain rightfully skeptical of any claims about that variable. Treating messages as random means that the deviant message affects both the main effect of treatment and the message x treatment interaction, so that treatment effects are evaluated with such message by message variations taken into account.

4. Methods are available (see Hedges & Olkin, 1985; Hunter, Schmidt, & Jackson, 1982) for studying the variability among the individual effect sizes, and some might suppose that this variability contains (or is) the message x treatment interaction. However, these methods apparently have a few unresolved defects (see Hedges & Olkin, 1986; Osburn, Callender, Greener, & Ashworth, 1983; Sackett, Harris, & Orr, 1986; Spector & Levine, 1987), and there is at this point little reason to prefer their results to direct assessments of the interaction.

REFERENCES

Bradac, J. J., & Mulac, A. (1984). A molecular view of powerful and powerless speech styles: Attributional consequences of specific language features and communicator intentions. *Communication Monographs, 51*, 307-319.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*, 335-359.

Clark, H. H. (1976). Reply to Wike and Church. *Journal of Verbal Learning and Verbal Behavior, 15*, 257-261.

Cody, M. J., Greene, J. O., Marston, P. J., O'Hair, H. D., Baaske, K. T., & Schneider, M. J. (1986). Situation perception and message strategy selection. In M. L. McLaughlin (Ed.), *Communication yearbook 9* (pp. 390-420). Beverly Hills, CA: Sage.

Coleman, E. B. (1964). Generalizing to a language population. Psychological Reports, 14, 219-226.

Cornfield, J., & Tukey, J. W. (1956). Average values of mean squares in factorials. *Annals of Mathematical Statistics, 27*, 907-949.

Doelger, J. A., Hewes, D. E., & Graham, M. (1986). Knowing when to "second-guess": The mindful analysis of messages. *Human Communication Research, 12*, 301-338.

Edgington, E. S. (1966). Statistical inference and nonrandom samples. *Psychological Bulletin, 66*, 485-487.

Ellis, D. G. (1982). Language and speech communication. In M. Burgoon (Ed.), *Communication yearbook 6* (pp. 34-62). Beverly Hills, CA: Sage.

Finch, P. D. (1976). The poverty of statisticism. In W. L. Harper & C. A. Hooker (Eds.), *Foundations of probability theory, statistical inference, and statistical theories of science: Foundations and philosophy of statistical inference* (Vol. 2, pp. 1-44). Dordrecht-Holland: D. Reidel.

Fontenelle, G. A., Phillips, A. P., & Lane, D. M. (1985). Generalizing across stimuli as well as subjects: A neglected aspect of external validity. *Journal of Applied Psychology, 70*, 101-107.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research.* Beverly Hills, CA: Sage.

Hample, D., & Dallinger, J. M. (1987). Individual differences in cognitive editing standards. *Human Communication Research, 14*, 123-144.

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

Hedges, L., & Olkin, I. (1986, October). Meta-analysis: A review and a new view. *Educational Researcher, 15*, 14-21.

Hewes, D. E., Graham, M. L., Doelger, J., & Pavitt, C. (1985). "Second-guessing": Message interpretation in social networks. *Human Communication Research, 11*, 299-334.

Housel, T. J. (1987). Conversational themes and attention focusing strategies: Predicting comprehension and recall. *Communication Quarterly, 33*, 236-253.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies*. Beverly Hills, CA: Sage.

Jackson, S., & Backus, D. (1982). Are compliance-gaining strategies dependent on situational variables? *Central States Speech Journal, 33*, 469-479.

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

Jackson, S., Jacobs, S., Burrell, N., & Allen, M. (1986). Characterizing ordinary argument: Substantive and methodological issues. *Journal of the American Forensic Association, 23*, 42-57.

Jackson, S., Jacobs, S., & Rossi, A. (1987). Conversational relevance: Three experiments on pragmatic connectedness in conversation. In M. L. McLaughlin (Ed.), *Communication yearbook 10* (pp. 323-347). Beverly Hills, CA: Sage.

Keppel, G. (1973). *Design and analysis: A researcher's handbook*. Englewood Cliffs, NJ: Prentice-Hall.

Keppel, G. (1976). Words as random variables. *Journal of Verbal Learning and Verbal Behavior, 15*, 263-265.

Keppel, G. (1982). *Design and analysis: A researcher's handbook* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Mulac, A., Bradac, J. J., & Mann, S. K. (1985). Male/female language differences and attributional consequences in children's television. *Human Communication Research, 11*, 481-506.

Mulac, A., Lundell, T. L., & Bradac, J. J. (1986). Male/female language differences and attributional consequences in a public speaking situation: Toward an explanation of the gender-linked language effect. *Communication Monographs, 53*, 115-129.

O'Keefe, B. J., & McCornack, S. A. (1987). Message design logic and message goal structure: Effects on perceptions of message quality in regulative communication situations. *Human Communication Research, 14*, 68-92.

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.

Planalp, S. (1985). Relational schemata: A test of alternative forms of relational knowledge as guides to communication. *Human Communication Research, 12*, 3-29.

Planalp, S., Graham, M., & Paulson, L. (1987). Cohesive devices in conversation. *Communication Monographs, 54*, 325-343.

Richter, M. L., & Seay, M. B. (1987). ANOVA designs with subjects and stimuli as random effects: Applications to prototype effects on recognition memory. *Journal of Personality and Social Psychology, 53*, 470-480.

Rogers, R. W., & Mewborn, C. R. (1976). Fear appeals and attitude change: Effects of a threat's noxiousness, probability of occurrence, and the efficacy of coping responses. *Journal of Personality and Social Psychology, 34*, 54-61.

Rosenthal, R. (1984). *Meta-analytic procedures for social research.* Beverly Hills, CA: Sage.

Sackett, P. R., Harris, M. M., & Orr, J. M. (1986). On seeking moderator variables in the meta-analysis of correlational data: A Monte Carlo investigation of statistical power and resistance to type I error. *Journal of Applied Psychology, 71,* 302-310.

Santa, J. L., Miller, J. J., & Shaw, M. L. (1979). Using quasi F to prevent alpha inflation due to stimulus variation. *Psychological Bulletin, 86,* 37-46.

Spector, P. E., & Levine, E. L. (1987). Meta-analysis for integrating study outcomes: A Monte Carlo study of its susceptibility to Type I and Type II errors. *Journal of Applied Psychology, 72,* 2-9.

Tracy, K. (1982). On getting the point: Distinguishing "issues" from "events," an aspect of conversational coherence. In M. Burgoon (Ed.), *Communication Yearbook 5* (pp. 279-301). New Brunswick, NJ: Transaction.

Tracy, K. (1983). The issue-event distinction: A rule of conversation and its scope condition. *Human Communication Research, 9,* 320-334.

Tracy, K. (1984). The effect of multiple goals on conversational relevance and topic shift. *Communication Monographs, 51,* 274-287.

Tracy, K., Craig, R. T., Smith, M., & Spisak, F. (1984). The discourse of requests: Assessments of a compliance-gaining approach. *Human Communication Research, 10,* 513-538.

Wickens, T. D., & Keppel, G. (1983). On the choice of design and of test statistic in the analysis of experiments with sampled materials. *Journal of Verbal Learning and Verbal Behavior, 22,* 296-309.