DOCUMENT RESUME

ED 310 452 CS 506 773

AUTHOR Slater, Michael D.
TITLE Messages as Experimental Stimuli: Design, Analysis, and Inference.
PUB DATE Aug 89
NOTE 23p.; Paper presented at the Annual Meeting of the Association for Education in Journalism and Mass Communication (72nd, Washington, DC, August 10-13, 1989).
Speeches/Conference Papers (150) -- Information Analyses (070)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Communication Research; *Data Interpretation; Higher Education; Literature Reviews; *Research Design; *Research Methodology; Research Problems
IDENTIFIERS *Message Stimuli

ABSTRACT

The use of messages as experimental stimuli brings with it problems regarding the interpretability and generalizability of findings. Some psychologists and communication researchers have argued that message stimuli must be treated as random effects. A review of the literature examined data regarding experimental designs used in recent experimentation in the communication field, and appraised alternative uses of message stimuli with respect to interpretability and generalizability. Also studied were researchers' assumptions implicit in the use of random or fixed effect models, as well as the appropriateness of each approach in certain contexts. Findings suggest the treatment of messages as random effects when a variable has been operationalized by selecting several messages representative of values on that variable, but not when the variable has been operationalized by manipulating the message or messages themselves. (Two tables of data are included, and 16 references are attached.) (SR)

Messages as Experimental Stimuli: Design, Analysis, and Inference

Michael D. Slater

Colorado State University

The author owes a considerable debt to Byron Reeves, Dept. of Communication, Stanford University; James zumBrunnen, Dept. of Statistics, Colorado State University; and Kevin Murphy, Dept. of Psychology, Colorado State University, for stimulating discussion and suggestions.

Michael Slater is an assistant professor in the Department of Technical Journalism, Colorado State University, Fort Collins, CO 80523, 303-491-5485.

For presentation to the annual conference of the Association for Education in Journalism and Mass Communication, Communication Theory and Methodology Division, Washington, D.C., August 1989.

Running head: MESSAGES AS EXPERIMENTAL STIMULI

Abstract

The use of messages as experimental stimuli brings with it problems regarding the interpretability and generalizability of findings. Some psychologists (e.g., Clark, 1973; Richter & Seay, 1987) and communication researchers (Jackson, O'Keefe, & Jacobs, 1988) have argued that message stimuli must be treated as random effects. This paper provides data regarding experimental designs used in recent experimentation in the communication field, reviews alternative uses of message stimuli with respect to interpretability and generalizability, discusses researchers' assumptions implicit in the use of random or fixed effect models, and suggests contexts in which each approach is appropriate. In particular, this paper argues that treatment of messages as random effects is appropriate when a variable has been operationalized by selecting several messages representative of values on that variable, but not when the variable has been operationalized by manipulating the message or messages themselves.

Messages as Experimental Stimuli: Design, Analysis, and Inference

Messages of various kinds--actual or constructed excerpts from publications, television and radio programming, conversation, or transcription of "real-life" events--are a principal source of stimuli for experimental research in many social science disciplines. In particular, message stimuli are of central interest in studies of mass communication and of persuasion. Implicit in the use of messages as stimuli, however, are internal and external validity problems that threaten the interpretability and generalizability of experimental message effects research. These problems include a) the difficulty of using any two different messages to represent different values on a single variable, given the many confounding differences between any two messages; b) the idiosyncratic nature of any given set of messages used in an experiment, making problematic the generalization of findings beyond the specific messages used in the experiment; c) the difficulty of defining a specific population of messages in order to make possible random sampling of messages for use in experimentation, and d) the logistic and methodological problems inherent in using large numbers of messages, in sampling messages that will be effective experimental manipulations, and in applying statistics that permit inference to a population of messages.

Both psychologists and communication researchers have debated strategies to address these problems. In particular, Clark (1973), Fontanelle, Phillips, & Lane (1985), Richter &

Seay (1987), and Jackson, O'Keefe, & Jacobs (1988) have argued for multiple-message designs, in which messages should be treated as random effects. They argue that the random effects model is needed in order to take into account variance due to messages selected as stimuli, and to make possible inference to larger populations. Morley (1988a, 1988b) has argued that such a strategy makes unrealistic statistical assumptions and poses unrealistic demands on the experimenter.

It may be useful to place the debate concerning treatment of messages as random versus fixed effects in a larger context. This paper will review a range of message experimentation strategies, suggesting consequences regarding interpretability and generalizability for each. In particular, assumptions about messages and about the intended scope of a given experiment implicit in random versus fixed effects analyses will be discussed. Criteria will be identified which distinguish between message experiments which do or do not require random effects analyses.

Single and multiple message stimuli. Using single messages to each represent different values of a variable--e.g., using two advertisements, one humorous, one not, in order to operationalize humor in advertisements--clearly tends to confound incidental message differences with differences associated with the variable. However, when the manipulation is within messages--that is, the same message is used to represent both values of the variable, with some single characteristic of the message manipulated--single message stimuli may be used effectively to develop theory.

For example, much of the Yale school of persuasion research was built up through studies using within-message manipulations: e.g., attributing the same message to different sources, or arranging the same message as first or last in a series of presentations (Hovland, Janis, & Kelley, 1953). Such two-stimulus, within-message designs apparently remain the mainstay of experimental communication research. Over half of the experiments using message stimuli reported in major communication journals over a period of a year employed this design (see Table 1). While within-message manipulation may strengthen the internal validity of single stimuli designs, external validity remains minimal.[1]

Table 1 about here

Jackson, O'Keefe, & Jacobs (1988) argue for using multiple messages to represent the same value on a variable. They suggest that message replication should mitigate systematic error due to the idiosyncracies of the messages representing that value. While such replication of messages obviously is preferable to the use of a single message, the fundamental problem of confounding has not changed. It is all too likely that systematic error will remain in the selection of messages. For example, any set of messages selected to represent a level of a variable such as humor will also tend to vary systematically along many associated dimensions: narrative style, visual complexity, use of dialogue and/or sound. It is hard to say whether the variable said to be manipulated--humor--is responsible for effects found, or if

some associated formal characteristic may be the source of effects (see Reeves, 1989, for a discussion of problems inherent in using message variables based on media distinctions such as news stories or advertisements, that are in fact composed of many more primitive components).

In addition, when one selects messages sharing some characteristic such as humor, there is a tendency to draw the messages from some superordinate category which may also confound findings. For example, a search for humorous advertisements may tend to draw from product categories distinctly different from advertisements that are somber in tone. These differences in product categories may be associated with attentional or cognitive processing differences that would be likely to influence experimental results (Petty & Cacioppo, 1986).

Despite these drawbacks, message replication remains a useful and relatively popular strategy for using messages to operationalize variables. Just under a third of the message stimuli experiments examined used more than one message to operationalize a given value on a variable (see Table 1). Several approaches may be used to compensate for the methodological weaknesses of message replication. One approach is reductionist: one may follow-up initial research efforts by systematically controlling and varying formal characteristics associated with the variable. In the example of humorous advertisements, one may seek to isolate the effects of visual complexity, cuts, etc., in determining their contribution to the overall effect.

A second approach is, as with single messages, to use within-message manipulations.[2] This was the approach used in five of the eight experiments examined which replicated messages (see Table 1). While within-message manipulations often alleviate the problem of confounds, they raise additional difficulties. While it is relatively easy to manipulate source cues within-message, other manipulations are much more problematic. For example, it would be difficult to remove the humor from a humorous advertisement without gutting the ad. It might be possible to select several messages made funny by jokes which may be excised. However, the experiment then speaks to the effects of jokes in messages, not to the question of humor generally.

A third approach would be to place the major emphasis on external validity and to sidestep some of these other internal validity problems. One might define a population of messages of interest, such as advertisements on primetime network TV over a three-month span, randomly select a sample of advertisements, and then use judges to sort the sampled messages on levels of humor. Then, one can argue that whatever humor is, those messages at different levels of perceived humorousness in the population produce certain predictable effects. Such a strategy raises additional methodological issues, which will be discussed below.

Messages: random vs. fixed effects. A fourth approach to dealing with the accidental confounding of messages and treatments, advocated by Clark (1973), Jackson & Jacobs (1983), Fontanelle, Phillips, & Lane (1985), Richter & Seay

(1987), and Jackson, O'Keefe, and Jacobs (1988), is statistical. They argue than messages should be treated statistically as a random effect in analyses of variance. Such a procedure, in their view, is necessary to "attempt to assess the magnitude of [accidental] confounding" (Wickens & Keppel, 1983:304, cited by Jackson, O'Keefe, & Jacobs, 1988).

This suggestion, as Morley (1988a) points out, has serious implications for research in communication. Random effects models are generally less powerful in detecting effects than are fixed effects models. Their use demands designs which use as many messages as possible to maximize power, placing heavy demands on experimenter ingenuity and resources. The position argued in this paper is that random effects models are not always necessary whenever message stimuli are used, contrary to the opinions of Fontanelle, Phillips, & Lane (1985), Richter & Seay (1987), and Jackson, O'Keefe, & Jacobs (1988). Nonetheless, random effects models may be necessary even in the absence of random sampling of message stimuli, contrary to Morley's (1988a) arguments. The appropriate analysis model, as will be detailed below, depends upon how variables are conceptualized and operationalized.

Generalization and message stimuli. Treating message stimuli as a random factor offers considerable potential for increasing the generalizability of message experimentation when messages are sampled randomly from a defined population. As Morley (1988a) points out, one can sometimes define a population of messages in such a way as to make random samples possible. For example, research questions may be posed in

terms of lead stories in major metropolitan newspapers in a given time frame, primetime network television advertisements in a three-month period, opening statements in university debate competitions, or summations by defense attorneys in county jury trials.[3] No such designs were found in the experiments examined (see Table 1).

If message stimuli are randomly sampled, using the techniques advocated by Clark (1977), Fontanelle, Phillips, & Lane (1985), Richter & Seay (1987), and Jackson, O'Keefe, & Jacobs (1988) make considerable sense. One could test effect size against the F or quasi-F ratio, and reasonably make a claim of inference to the larger population.[4] Such experiments would be unusually persuasive with respect to external validity, especially when a large sample of messages is used. Richter & Seay (1987) and Jackson, O'Keefe, & Jacobs (1988) recommend designs confounding subjects and messages--in which no two subjects receive the same message--as one means to maximize the size of the message sample. Jackson, O'Keefe, & Jacobs (1988) point out that using such confounded designs, or using nested designs, avoids the need for the approximated quasi-F ratios. The major drawbacks of random sampling, of course, include the impossibility in many cases of defining populations accessble to sampling, the logistical demands of locating and creating experiments with often lengthly natural messages, and the possible weakness of experimental manipulations resulting from use of randomly rather than purposively selected stimuli.

In the absence of random sampling of messages, the

question of random versus fixed effects models is more complex. Conceptually, a factor should be treated as random when the variable consists of many possible replications, and some of those replications are selected in order to operationalize the variable (Winer, 1971). As Clark (1973) and Jackson, O'Keefe, & Jacobs (1988) point out, selection of replications need not be random; purposive selection of replications also requires use of the random effects model.

Treating messages as random effects, then, is clearly the appropriate analytic approach when a message selection procedure is used in order to operationalize a variable. This is commonly the case in between-message manipulations, in which prototypical messages of different types are selected and opposed to one another as different treatment conditions. For example, one may operationalize fiction versus non-fiction by selecting samples of fiction and non-fiction. Then, the various messages represent a purposive sample of the variable, and should be treated as random. However, this evidentally has not been common practice in communication research (see Table 1).

It is important to remain aware that treatment of messages as random in this case is simply a necessity borne of the assumptions of the ANOVA model. Little is gained in terms of generalizability--significance of a random factor in the absence of random sampling does not permit statistical inference to a meaningful population of messages. Nor is anything truly gained in terms of control over accidental confounding of messages and treatments, as one gains through

within-message manipulations, or designs that permit partitioning of variance due to message differences out of the error term (e.g. Greco-Latin and related designs). The necessity of using the random effects model in this case means both a probable loss of power and increased demands on the experimenter in terms of using multiple message stimuli.

The position taken in this paper, however, differs from that of Fontanelle, Phillips, & Lane (1985), Richter & Seay (1987), and Jackson, O'Keefe, & Jacobs (1988) in one crucial respect. These authors argue that whenever messages are used in an experiment, they are implicitly sampled, and random effects models should be used. In our view, random effects models are inappropriate when within-message manipulations are used.

Why are within-message manipulations different from between-message manipulations with respect to random versus fixed effects? In the within-message manipulation, the variable has been not been operationalized using either a purposive or random sampling procedure. For example, if the same message (or set of messages) is labelled variously as fiction and non-fiction when presented to alternate subjects, it is the label and not a sampling procedure that is used to operationalize the fiction-non/fiction variable. An excerpt taken from a novel and labelled as non-fiction is clearly not a sample of non-fiction. Rather, it represents a single demonstration of the label change's effects.

Messages when used as the basis for manipulations are conceptually no different from the situations constructed and

manipulated by social psychologists. If one were to treat within-message manipulations as random effects, one would also have to do the same for constructed social situations. Consider Milgram's (1963) experiments, in which subjects gave electric shocks they believed to be dangerous to experimental confederates, on the instruction of the experimenter. One would, if one used the same logic of treating manipulated messages as random effects, have to consider Milgram's situation as being sampled from some imaginary population of constructed authoritarian situations, and treat such situations as random effects. One might argue that social psychological experimental situations are fabricated, and many messages used as the basis for manipulations are excerpted from natural messages. However, once those messages have been experimentally manipulated, they no longer are part of a population of natural messages. Moreover, as in the social psychological experiment, it is the manipulation--and not the selection of messages--that is used to operationalize the variables studied.

This argument, of course, holds only for within-message manipulations. One might argue that between-message manipulations also may be treated as fixed effects if no attempt is made at inference to messages other than those used in the experiment. Such an approach, however, is logically inconsistent. Between-message manipulations usually consist of opposing messages selected in order to represent different values on a variable: fiction vs. nonfiction, or humorous vs. non-humorous. If messages are treated as fixed,

then the messages in the experiment can be taken to represent only themselves. They no longer serve as exemplars of the variable. The variable--in this example, fiction vs. nonfiction or humorous vs. non-humorous--then cannot truly be said to have been manipulated. One would be hard put to interpret results of such an experiment meaningfully, given the lack of a conceptually clear experimental manipulation.

The use of the fixed effects model for within-message manipulations, then, trades inferential power for power to detect treatment effects: Generalization cannot be made on statistical grounds to messages other than those used in the experiment. The central logic to such study is to attempt to demonstrate the existence of phenomena predicted by theory, not to provide evidence of generality. Generalization must proceed incrementally, through theoretical exploration and additional research.

From this perspective, between-message designs in which message stimuli have been sampled purposively do have one advantage over within-message designs. In both cases generalization is possible only through theoretical argument, not through statistical inference. However, the obstacle to generalization in the between-message design is the questionable representativeness of the messages selected as stimuli. The within-message design not only is confronted with the sampling problem, but must address the consequences of the artificiality of the manipulation used, which removes the stimuli a step further from nature.

These arguments can be presented in terms of a 2x2

matrix. One axis concerns how the variable is operationalized, and whether it demands fixed (within message manipulation) or random (between message manipulation) treatment. The other axis concerns whether or not the experiment is intended to permit statistical inference to a population for clearcut claims of generality.

Table 2 about here

The cell in the upper left is straightforward: if the variable has been operationalized through manipulation and not selection, one has a fixed effects model with no statistical basis for generalization (this of course does not exclude theoretical bases for generalization). The lower left cell represents the case in which messages have been purposively sampled: one has a random effects model, again with no statistical basis for generalization to a population. In the lower right cell, one has sampled messages randomly from a defined population: one has a random effects model permitting inference to that population. In the upper right, one has an impossible model. If one manipulates the message, that message is no longer representative of any population in the real world. It would not be meaningful to treat it as a random effect, or to attempt to make inference to a population. For example, if the experimenter randomly samples fiction and non-fiction messages, and then creates a second set of messages by switching the fiction/non-fiction labels, she may have provided some control over differences in message content. However, since half of her "fiction" messages are

now non-fiction messages with a false label, attempting to draw statistical inference to the population of fiction messages would be specious.

Conclusions. Several general conclusions, then, may be drawn. Treating message stimuli as the source of random effects can be a powerful tool in increasing generalizability when the research question is posed in terms of a defined and accessible population of messages, and when a relatively large number of stimuli are used. The inferential power of experiments is constrained by the inability, in most cases, to draw subjects randomly from a defined population. However, the prospect exists for inferentially powerful experiments using randomly sampled message stimuli. Even in such cases, though, the need remains to explore whether or not findings generalize further to populations of messages other than the one defined.

When the research question is not be posed in terms of a defined population of messages, then a design should be selected that minimizes the risk of confounding message and treatment effects. The appropriate design, of course, will depend upon the research question and the type of message or medium studied. Within-message manipulations offer superior internal validity, but some variables are difficult to manipulate within-message without introducing possible confounds (audio-visual media sometimes pose special within-message manipulation problems). The Greco-Latin square and related designs can partition variance due to message differences and increase power to detect treatment effects.

However, purposive selection of message stimuli to represent variables, without use of within-message manipulations, are more vulnerable to accidental confounding of message differences and treatments and, contrary to much current practice, necessitate the use of less powerful random effects statistical models. They therefore represent a normally less desirable alternative. Under some logistic constraints, however, such designs may be the only ones feasible. In such cases, replication remains the only protection against confounding message differences and treatments. Possible systematic differences between the kinds of messages used to operationalize different treatment conditions should be carefully considered prior to message sele ion, and carefully reviewed in discussions of findings.

Any single research effort--and experimental research in social science in particular--is limited in scope. As Fontenelle, Phillips, & Lane (1985) point out, too often researchers claim unreasonable generality with respect to messages. A valuable service is rendered if researchers acknowledge that, in the absence of defining and sampling message stimuli from a population, findings apply with certainty only to the messages at hand. Such an acknowledgement forces definition on theoretical grounds of the likely generality of findings.

Discussions of generality should focus on two topics: the possibility of accidental confounds which may, entirely or in part, explain findings; and the extent to which findings are likely to apply to various domains of messages other than

those used in the experiment. One researcher's accidental confound is another researcher's independent variable. Can we identify and propose variables which would lead to different results, different relationships? Do these relationships vary in predictable ways among various types of messages? It is this sort of questioning which, in the long run, advances the field.

References

Calfee, R.C. (1975). Experimental methods in psychology. NY: Holt, Rinehart, & Winston.

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

Glass, G.V., & Hopkins, K.D. (1984). Statistical methods in education and psychology (2nd edition). Englewood Cliffs, NJ: Prentice-Hall.

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.

Hovland, C.I., Janis, I.L., & Kelley, H.H. (1953). Communication and persuasion. New Haven: Yale University Press.

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

Jackson, S., O'Keefe, D.J., & Jacobs, S. (1988). The search for reliable generalization about messages: a comparison of research strategies. Human Communication Research 15 127-142.

Milgram, S. (1963). Behavioral study of obedience. Journal of Abnormal and Social Psychology 67. 371-378.

Morley, D.D. (1988a). Meta-analytic techniques: when generalizing to message populations is not possible. Human Communication Research 15 112-126.

Morley, D.D. (1988b). Reply to Jackson, O'Keefe, and Jacobs. Human Communication Research 15 143-149.

Petty, R.E., & Cacioppo, J.T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 19). NY: Academic Press.

Reeves, B. (1989). Theories about news and theories about cognition: arguments for a more radical separation. American Behavioral Scientist, July issue.

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 (3) 470-480.

Slater, M.D. (1989). Relative familiarity of the social group in a message as a determinant of the effects of fiction and non-fiction messages on subsequent beliefs. Manuscript under review.

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

Winer, B.J. (1971). Statistical principles in experimental design (2nd edition). NY: McGraw-Hill.

Notes

1. Some risk of confounding remains even with skillful within-message manipulations. As Jackson, O'Keefe, & Jacobs (1988) point out, even single-word changes of a message may introduce unintended sources of variation beyond the variable meant to be manipulated.

2. The Greco-Latin square and related designs have interesting possibilities in this regard, when experimental manipulations and intended analyses permit their use, as they provide a means for partitioning both subject variance and variance due to the original messages used as a basis for the experimental manipulations (Calfee, 1975; Slater, 1989; Winer, 1971). This removes important sources of error from the error term and increases experimental power.

3. A caution is in order here. As Richter & Seay (1987) point out, sampling from a population of finite size violates an ANOVA assumption: selection of one stimulus from the population increases the probability of selection of subsequent stimuli, so that probability for selection is not independent. The size of the message population defined, therefore, should be large enough to make the difference in probability trivial.

4. Fontanelle, Phillips, & Lane (1985) and Richter & Seay (1987) provide examples detailing how to calculate the pseudo-F ratio when messages are treated as random effects.

Table 1. Treatment of message stimuli in experiments reported in six communication journals during one year.

	HCR	JA	JBEM	JC	CR	JQ	Total	%
Single message:								
within[a]	6	5	2	1	1	0	15	58
between	0	0	0	0	0	2	2	8
Replicated messages:								
within	0	0	0	1	1	3	5	19
between	1	2	0	0	0	0	3	12
Subjects/ messages confounded:								
within	0	0	0	0	1	0	1	4
between	0	0	0	0	0	0	0	0
Nested or Greco-Latin:								
within	0	0	0	0	0	0	0	0
between	0	0	0	0	0	0	0	0
Random effects:								
within	0	0	0	0	0	0	0	0
between	0	0	0	0	0	0	0	0
Total:	7	7	2	2	3	5	26	

[a]Single message per treatment, with manipulations being within-message, include several cases in which only a single, unvaried message was used in the experiment. Instead, treatment conditions varied among subjects. Also, in several cases manipulation of treatments was considered to be within a single message even though messages differed in more than minimal ways, if there was an obvious attempt to use fundamentally the same message in both treatments.

Note. All available 1988 issues of Human Communication Research (HCR), Journal of Advertising (JA), Journal of Broadcasting and Electronic Media (JBEM), Journal of Communication (JC), Communication Research (CR), and Journalism Quarterly (JQ) were included in the analysis. If 1988 volumes were not complete at the time of data collection, an equivalent number of issues from the latter part of 1987 were used to represent a single volume.

Table 2. Alternative approaches to design and analysis of message experiments.

Type of experimental manipulation	Type of statistical inference intended: Stimuli only	Type of statistical inference intended: To population
Within message (fixed effect)	Standard fixed effects model experiment	Impossible
Between message (random effect)	Stimuli are random effect, no special generalizability	Stimuli are random effect, generalizes to population randomly sampled