Experimental Methodology in Journalism and Mass Communication Research

Article *in* Journalism & Mass Communication Quarterly · March 2012

DOI: 10.1177/1077699011430066

3 authors:

Esther Thorson
Michigan State University

Robert H. Wicks
University of Arkansas

Glenn Leshner
University of Oklahoma

Journalism & Mass Communication Quarterly
89(1) 112–124
© 2012 AEJMC
Reprints and permission: http://www.sagepub.com/journalsPermissions.nav
DOI: 10.1177/1077699011430066
http://jmcq.sagepub.com

Experimental Methodology in Journalism and Mass Communication Research

Esther Thorson[1], Rob Wicks[2], and Glenn Leshner[1]

Abstract

Experiments are a powerful method for understanding causal relationships in journalism and mass communication research. In this essay, the authors examine seven aspects of experimental quality that reviewers should include as criteria in their evaluations. They note that there are complex interrelationships among these indicators. In cases where aspects of the standards are controversial, the authors attempt to summarize the conflicting arguments. Where different methodological conclusions can be rationalized as appropriate, the authors' suggestion is that the researcher make clear what decisions were made in the experimental design and why, so that readers can evaluate those decisions.

Keywords

experiments, control, publication criteria

Experiments are important to the theoretical development of fields like journalism and mass communication because they provide the most rigorous way to establish causal relationships between independent and dependent variables (as well as moderators and mediators), relationships critical for building and evaluating theory. Experimentation is not a dominant approach in *Journalism & Mass Communication Quarterly*. Indeed only 12% of social-scientific manuscripts submitted to the journal in 2008–9 were based on experiments.[1] Grabe and Westley reported only 9% to 13% of studies in mass communication use experimental methods.[2] Nevertheless it is important that the experiments that do appear are of the same quality as they are in fields where they play

[1]University of Missouri–Columbia, Columbia, MO, USA
[2]University of Arkansas–Fayetteville, Fayetteville, AR, USA

Corresponding Author:
Rob Wicks, University of Arkansas–Fayetteville, Department of Communication, Kimpel Hall, Fayetteville, AR 72701
Email: rwicks@uark.edu

a more dominant methodological role. In this essay we identify common threats to standards of experimental research quality. We do not provide a how-to summary for all experimentation, which is widely available in experimental design texts. Rather, we focus on issues that are often complex and even controversial.

Specifically, we examine seven attributes of well-executed experiments. They all involve situations where threats to the quality of social science experiments commonly occur. Conflicting arguments are presented as a means of explaining the complex interrelationships that exist among these indicators. Our objective is to explicate the reasons why these debates exist and to provide guidance to both authors and reviewers so that reasoned decisions can be made during the process of developing publishable work in the field. For the sophisticated student of experimental methods, there are many additional complexities to deal with. The references included here will also provide some suggestions for further study.

Seven Attributes of Well-Executed Experiments

1. Explication of the theory being tested and explanation of how the posited relations among independent, dependent, moderator, mediator, and control variables relate to that theory
2. Explication of how the experimental design will demonstrate causal relations between independent and dependent variables
3. Clarity in conceptualizing media stimuli
4. Clear identification of hypotheses and research questions
5. Clear specification of the sample and acknowledgment of its limitations
6. Correct specification of effect size, power, number of participants, and alpha levels
7. Consideration and empirical assessment of alternative explanations of experimental findings such as the following:

   a. Confounds
   b. Message-related variance
   c. Manipulation checks and/or pilot testing
   d. Counterbalancing or randomization of presentation orders and conditions

1. Explication of the Theory Being Tested and Explanation of How the Posited Relations Among Independent, Dependent, Moderator, Mediator, And Control Variables Relate to That Theory

Reviewers should check that posited independent–dependent variable links (including moderators, mediators, and controls, if applicable) are rationalized within a compelling theory. A central problem for the communication field is that there are often many theoretical approaches that can prove useful. For example, Potter points out there are at least ten theories that have been used to explain why aggressive behavior shown in messages like television programs and movies would cause aggressive behavior in young viewers (e.g., theories like excitation transfer, catharsis, social learning, priming, cultivation, etc.).[3] In reports of experiments it is necessary to explicate what theory is being tested and then ensure that all hypotheses tested in the study link to that theory. Sometimes competing theories may predict different outcomes that can be juxtaposed in hypothesis form. Authors should always make clear how the results of an experiment reflect back on theory or competing theories.

2. Explication of How the Experimental Design Will Demonstrate Causal Relations between Independent and Dependent Variables

Reviewers should look for clear statements about how experimental conditions will be used to show that independent variables actually affect dependent variables. For example, to establish a causal relationship between exposure to mediated aggressive behavior (e.g., actions depicted in video clips) and aggressive behavior by children exposed to the video clips, the researcher must demonstrate that when children are exposed to the mediated aggressive behavior, they exhibit similar aggressive behavior to that depicted, and when they are exposed to mediated content that does not contain aggressive behavior, they do not exhibit the aggressive behavior. Video content without aggressive behavior provides the control stimulus against which the experimental condition is compared. If children have been randomly assigned to the two conditions, then it can be said that the aggressive content causes aggressive behavior. However, without other variables that serve as mediators or moderators, it will not be possible to say why the aggressive behaviors in the video clips caused the children's aggressive behaviors (e.g., that they were aroused by the behaviors show in the video clips).

3. Clarity in Conceptualizing Media Stimuli

Reviewers should look for theoretical and operational clarity about how media stimuli are defined. Especially when the goal of an experiment is to link psychological responses to physical attributes of media stimuli, manipulations of those physical attributes of the stimuli are hypothesized to cause changes in the dependent variable (e.g., the number of violent acts in videos would be an example of a physical message variable).

There is controversy about how best to characterize the structure of media stimuli. For example, many scholars treat media stimuli in "industry units" (e.g., commercials, news stories). Others suggest it is more useful to describe them in terms of variables more closely related to psychological processing (e.g., visual complexity, brightness, contrast, movement of objects on the screen).[4] Tao and Bucy argue that media stimuli named in terms of posited psychological impact (e.g., aggressive television content) are problematic because they fail to specify the physical stimuli attributes that cause the psychological response (e.g., What content characteristics are indicative of aggression?), thus conflating the physical stimulus with the psychological response.[5] The bottom line is that it is important to select the physical stimulus features considered important in the tested theory, treat them as the independent variables, and then measure psychological responses, which may include mediators or moderators as well as dependent variables.

Sundar provides an enlightening example. He points out that if interactivity is conceptualized as a feature of the physical stimulus, then it is a mistake to operationalize it in terms of people's perception of interactivity.[6] Instead, it is necessary to develop a theory of what in a stimulus makes it interactive (e.g., features like functionality, user accommodations, organization, control, choice, and contingencies). These features are physical, not psychologically perceived independent variables.

4. Clear Identification of Hypotheses and Research Questions

Reviewers should look for clear statements of hypotheses and research questions. Predictions about how independent variables are expected to be related to dependent variables are generally provided in the form of either a directional hypothesis or a research question. The distinction between the two depends on how specific one's theory is and/or how much prior evidence is available. If a number of relevant prior studies have found or suggested a specific direction of independent–dependent variable relationships, or if the tested theory leads to specifically deduced predictions, hypotheses are used to state the posited relationship between independent and dependent variables. If neither theory nor prior research leads to specific predictions about the relationships between independent and dependent variables, research questions should be used. Hypotheses derived from theory will posit not just "differences" between conditions but directionality of the differences (e.g., there will be more aggressive behavior when children watch longer programs depicting violent events than when they watch shorter programs depicting violent events).

It should be noted that hypotheses are more powerful analytic tools in experiments than research questions because they serve as bridges between theory and observation. Some argue that hypotheses should be stated in terms of relationships between concepts, not between operationalizations of the variables.[7] The rationale for this is that the concepts are directly represented in the theory, and the theory provides explanation of why they should be linked. Operationalizations are designed to be physical measurements of concepts but may or may not capture theoretical concepts successfully. It is useful to distinguish theoretical (conceptual) statements of hypotheses and operational (experimental) statements of hypotheses making clear the linkage between them. For example, in a study of the impact of direct and indirect experiences with a product,[8] the stated hypothesis was, "Direct product experience will trigger a more concrete mental construal than an indirect product experience." Direct product experience, a theoretical variable, was operationalized as hands-on use of the product (an MP3 player), indirect experience (also a theoretical variable) as having information presented about the player from a PowerPoint. Mental construal (the third theoretical variable) was operationalized in terms of open-ended participant-generated descriptions of "using the product" (e.g., how to turn it on, loading music onto it, changing the volume). Thus, it was clear that the theoretical hypothesis was evaluated with a test of the operational hypothesis.

Hypotheses can propose different relationships between independent and dependent variables: direct effects, indirect effects (either mediated or moderated), or a combination of these. Sometimes hypotheses are used to describe how control variables are expected to influence dependent variables (e.g., individual difference variables like gender, income, education can be measured and treated as controls before examination of the influence of the independent variable). It is important that the author makes clear just what kinds of relations are hypothesized.

Smith, Levine, Lachland, and Fediuk argue that when it comes to hypotheses, fewer variables are better than more.[9] They point out that if every dependent variable is related to every independent variable, then the number of hypotheses will be at least the number of independent variables multiplied by the number of dependent variables. The number of statistical tests applied to test those hypotheses will then increase the likelihood of a Type I error, that is, the likelihood of claiming support for relationships that do not exist. Of course, the calculation of the number of hypotheses does not take into account mediating and moderating relationships, which will further increase the number of statistical tests.

5. Clear Specification of the Sample and Acknowledgment of Its Limitations

There has been considerable discussion of whether it is necessary to randomly sample experimental participants from a targeted population.[10] When researchers randomly sample from a population, they can infer from findings in the sample to a particular population. For example, if a researcher obtains a random sample of people in a county with phone lines in their home, then the mean number of those sampled who also have cell phones can be used as a parameter estimate of how many people in the county (the population) have cell phones as well as land lines. And the percentage of people who respond positively (dependent variable) to an offer of broadband at a particular price (independent variable) can be used to estimate the relationship between that offer and a positive response in the county population. Random sampling thus enables statistical generalization from sample features to population features.

Samples for experiments, however, rarely involve random samples. In fact, one does not have to read many reports of experiments in mass communication to recognize the fact that nearly none employs samples randomly selected from populations. Instead, experimental researchers typically acquire convenience samples (like second graders from several school districts in town, college students enrolled in large communication classes, or adults who agree to participate in an experiment for a chance to win a digital music player). These individuals are then randomly assigned to the conditions in the experiment. Because there is no random sampling of participants, inferences cannot be applied to the likelihood that values found in the experiment are representative of values that would be found in the population as a whole.[11] Instead, logical inferences are made about the multivariate relationships among the variables in the experiment.[12] The "population" in this case is all possible samples one could randomly draw from the group of individuals in the experiment.

The strength of those logical inferences depends on how well the experimenter can make the argument for them. Suppose a study looks at psychophysiological processes that are unlikely to be affected by demographics or sociocultural variables. Then generalizing to "human processing" could be argued since those processes are expected to be similar across other samples.[13] On the other hand, if testing involves variables clearly related to socially learned differences like attitudes about social mores, then generalization beyond people who do not share those mores would be unconvincing. Often the generalizability of experimental findings is also supported by replication of the experiment across different groups in different contexts. A good example of a finding that seems to be robust regardless of when or where it is tested is the third-person effect.[14]

It should be recognized, however, that sometimes similarity of independent variable–dependent variable links across quite different populations is surprising. For example, Basil and his colleagues conducted a study of celebrity effects around the time of the death of Diana, Princess of Wales, with three different samples: a random telephone sample drawn from seven states, a nonrandom sample of college students in three states, and a nonrandom web-based survey sample. The samples varied widely in terms of age and gender, yet Basil and his colleagues found a remarkably consistent relationship across the different samples between respondents' identification with Diana and their media use.[15]

In all experimental reports, sample characteristics and selection methods should be included so the reader can evaluate all claims to generalizability.

6. Correct Specification of Effect Size, Power, Number of Participants, and Alpha Levels

Reviewers should look for appropriate specification of effect size, power, number of participants, and alpha. There are complex interrelationships among effect size, the power of a study, the size of the effects sought or found, the number of participants tested in an experiment, and the statistical criterion chosen for rejecting the null hypothesis. The power of a study refers to the probability of the study to detect a certain size effect. A power analysis should be conducted prior to executing a study to determine the number of participants the experimenter should include. To compute an a priori power analysis, the research design must be specified (e.g., the number of between-subjects factors, the number of repeated measures, the correlation among repeated measures), the type of statistical analyses to be conducted (e.g., ANOVA, regression), Type I error rate (α; convention sets it a .05), and the size of effect sought. The a priori effect size is generally estimated by either prior literature or "rules of thumb."[16] Power tables can be used to compute an a priori power analysis, as can computer software.[17]

Too often reviewers criticize the number of participants in a study as being too small without consideration of the study's power. Sometimes, a study's sample is so large that there will be relationships found among all variables. In these cases, the effect size that is detected and inferred as "statistically significant" may be so small that the effect is theoretically uninteresting and/or trivial. Hence, a larger sample size is not always better. The reviewer should be familiar enough with the concept and operation of power so that sample size, effect size, observed power, and levels of statistical significance can be correctly evaluated.

Editors and reviewers should require the reporting of effect sizes in addition to the reporting of statistical significance levels. Effect size can help the reader determine whether the observed relationship among variables is valuable.[18] Statistical software packages include various effect size statistics, some of which may not be the most appropriate for a particular analysis. For example, SPSS provides partial η^2 in ANOVA output when η^2 may be more appropriate.[19] Many statistics textbooks and articles include formulas that permit the hand calculations of these statistics.[20]

7. Consideration and Empirical Assessment of Alternative Explanations of Experimental Findings

As we have seen, a significant challenge to the validity of experimental findings is the possibility that the independent variables did not really cause the observed changes in the dependent variable, but rather the effect was the result of some unrecognized source of influence.

Reviewers should consider at least four issues that can threaten the validity of experimental findings: confounds, message-related variance, lack of manipulation checks, and order effects.

a. Confounds. One highly problematic threat to the quality of experiments is confounding variables. Confounds are sources of variation that are inextricably linked with the independent variable and occur because of errors in operationalizing an independent variable. For example, suppose your independent variable is reading material difficulty, operationalized as vocabulary sophistication. Further suppose that the difficult material is longer than the easy material. In this case, vocabulary sophistication is confounded with length. This would mean it would be impossible to attribute any observed difference in the dependent variable to either unique vocabulary effects or length effects. Confounds generally are fatal flaws for experimental research, and therefore reviewers should look carefully for their presence.

b. Message-related variance. *Reviewers should look for* appropriate treatment of messages tested in experiments. A significant issue, especially in research that assesses psychological responses to media messages,[21] has to do with how researchers create variance in independent variables. The arguments presented here are informed by similar arguments found elsewhere.[22]

There are two kinds of variance that are important in experiments: treatment variance and message variance. When researchers are interested in how people respond to media messages, treatment variance refers to the message manipulations, that is, how the levels of each independent variable vary in a message. Too often, experimenters focus primarily on treatment variance and not enough on message variance.

For example, if we were interested in how news sources affect perceived credibility of a story, we might choose a news story and create two versions: one in which the primary source is the government and another in which the primary source is a private citizen. In this case we create the variance of our two levels of treatment (government source vs. private citizen) and hold all other message features the same. Creating treatment variance through message alteration is common in our field.

The issue with this design is that it ignores message variance. Message variance refers to employing multiple messages per treatment level, which strengthens the ability to generalize to message categories. We create message variance because we rarely are interested in one particular message. Rather, we normally are interested in *kinds* of messages, or message categories to which our manipulation would apply. When we forgo message variance in an experiment, we have opted for a single-message design.

Single-message designs occur when an experiment includes only a single message to represent a treatment level. Even if treatment variance is created via message alteration, the problem with this design is that any conclusion that can be made about the effect of the manipulation must be constrained to that particular message. We cannot conclude that the *type* of message had an effect, but only the exact message that was used had an effect. Since we rarely theorize about one message, it is difficult to imagine how such a design would be a useful contribution to theory.

What would one conclude if any effect were found that was based on a single message? Since media messages vary widely, the feature of interest will likely be confounded with a multitude of other message features that are concurrently manipulated, of which one does not (and perhaps could not) know. Any of these features might moderate the relationship between the manipulation and the dependent variable. The results might show a main effect of the independent variable on the dependent variable, but that apparent relationship might be the result of an interaction with another, unknown message factor. The problem is that it is nearly impossible to account for all the possible confounds by using a single message.

Another strategy for creating treatment variance is to employ a sample of messages, such that each level of an independent variable is represented by several messages. When we select a sample of messages to represent each treatment level, we would ideally sample from a knowable message population that contains all of the messages that represent a particular treatment. Rarely, however, do we have access to such a population of messages. The best we can do is find multiple messages that represent one and only one treatment, where the variance within treatment ideally would be less than the variance between treatments. Given that messages vary on a host of features other than the ones of interest, many of which we cannot know, the best strategy here is to include as many messages per treatment level as we can reasonably expect a participant to attend to without fatigue or boredom setting in. The more messages naturally vary within a treatment, the larger the number of messages is needed. This strategy creates both treatment and message variance.

For example, Leshner, Bolls, and Thomas asked if antitobacco messages that contained both a health threat and a negative graphic image were more successful at engaging cognitive and emotional processes than antitobacco messages that contained either only one or neither of these features.[23] They selected twenty-four messages from a collection available to them to represent one of four treatment levels (six messages per treatment): health threat–negative image, no health threat–negative image, health threat–no negative image, and no health threat–no negative image. Clearly, the messages within a single treatment vary widely, but the idea is that the within-treatment message variance is treated as random error in tests between treatments. The significant findings of their study occurred despite within-treatment message variance.

Another option of creating both treatment and message variance is to employ both message alteration and message repetition strategies. An example of a study that employed both was a 3 × 2 experiment, where the researcher manipulated story type (three levels: live, breaking, traditional) and emotional message content (two levels: threat, negative graphic images).[24] In this study, the researcher altered messages to create treatment variance for story type. She created three versions of each story by having an anchor verbally refer to a story as "live" or as "breaking," or by including no verbal designation for the "traditional" story. To create treatment variance for emotional content, three stories were chosen to represent the threat condition and three different stories were chosen to represent the negative graphic image condition. The researcher employed a fractional design, such that each participant saw six stories that represented the six levels of the two independent variables (traditional–threat, traditional–graphic image, live–threat, live–graphic image, breaking–threat, breaking–graphic image, presented in random orders). No participant saw two versions of the same story. By including a message repetition factor for the emotional content treatment, the researcher was better able to generalize to other similar emotional news stories than if she had relied on only one story per condition.

Although message repetitions may require more work on the part of the researcher, their inclusion in an experimental research design greatly strengthens what can be learned from a study.

c. Manipulation checks and/or pilot testing. Reviewers should look for evidence that before a group of messages appear in an experiment they have been pretested in terms of their generation of the posited psychological responses. For example, if the conceptual hypothesis is news stories that frame poverty as resulting from individual rather than social responsibility make people less likely to support increasing unemployment benefits, then a pretest of stories with each frame (individual or social responsibility) should be conducted to determine whether each one is perceived consistently with its "frame."

On the other hand, what were once considered "manipulation checks" of the impact of stimulus attributes may best be treated as measures of mediating variables between the physical attributes of the stimuli and the dependent variable(s) of interest. Suppose an experiment is conducted to determine whether the number of anonymous sources reduces the credibility of newspaper stories. In the experiment, participants read four stories, each with five anonymous sources and four stories with no anonymous sources. Should there be measurement of the perception of the number of anonymous sources in the stories?[25]

We suggest this depends on the theory being tested. If the theory posits that people are sensitive to the presence of anonymous sources and if the presence of anonymous sources reduces people's perception of the truth of what is claimed, then it is important to establish that the presence or absence of anonymous sources is perceived. When the theory provides this explanation, perception of the number of anonymous sources becomes a mediator (rather than a manipulation check) and should be measured. O'Keefe points out that this means most of what have been referred to in such studies as "manipulation checks" are best conceptualized as mediating variables.[26]

d. Counterbalancing or randomization of presentation orders and conditions. Reviewers should look for appropriate treatment of the orders of stimulus conditions and questions. The treatment orders in which experimental participants are presented with stimuli and the order of questions participants must answer always make a difference. Order effects can be caused by such variables as opportunity for differential practice, by fatigue, or by the impact of responses to the prior stimulus on a subsequent stimulus (i.e., carryover effects). It is therefore crucial to employ either randomized or counterbalanced orders in experiments. Of course, sometimes there is a logically required order of questions. For example, free recall must precede cued recall, which must precede recognition.[27] The experimental design should pay careful attention to any source of order effects—on treatment orders, in question orders, even in the order of times of day during which between-subject conditions are executed. The handling of order effects should be clearly explained in the research report.

If the stimulus materials for the experiment are computer based, randomization is usually convenient and helps ensure that whatever effects there are of getting condition or question A before or after B will be equally distributed across participants. When different packets of stimulus materials must be prepared, it is often more convenient to use some counterbalanced design for the orders of the experimental stimuli and tasks.

Concluding Thoughts

We attempted to present seven aspects of experiments that reviewers should look for when evaluating experimental research in mass communication. The intent is not necessarily to prescribe particular ways of doing things, but rather to illustrate some issues that sometimes are neglected or misunderstood. We also attempted to highlight important discussions in the literature where issues may not be settled. In any case, we hope that reviewers and researchers carefully consider the issues raised here as experimental research is designed, conducted, and evaluated.

Certainly there are many important issues of experimental research that we did not address. We decided not to discuss issues of design, analysis, or ethics, among others. We reasoned that such issues are dealt with adequately in research methods texts, journal articles, and book chapters. We do not discount the importance of such issues; rather, we chose to focus on issues that may not be widely recognized or shared among researchers in our field.

An experiment is a powerful method for getting at causal relationships, but as can be inferred from this description of the challenges of the method, a single experiment in isolation usually allows for many interpretations. In fact, if one looks at a journal that focuses on systematic evaluation of alternative theories, such as *Psychological Review*, one sees that experiment-based articles frequently involve multiple experiments, each one of which progressively eliminates a single alternative interpretation. Designing experimental conditions, designing stimuli, and testing participants are difficult and time-consuming. Nevertheless, findings from high-quality experiments can be field changers. The recommendations articulated can be supplemented with the recommended readings listed in the appendix.

Appendix

Suggested Readings

American Psychological Association. *Publication Manual of the American Psychological Association*. 6th ed. Washington, DC: American Psychological Association, 2010.

Calfee, Robert C. *Experimental Methods in Psychology*. New York: Holt, Rinehart & Winston, 1985.

Campbell, Donald T., and Julian C. Stanley. *Experimental and Quasi-Experimental Designs for Research*. Boston: Houghton Mifflin, 1963.

Cohen, Jacob. “A Power Primer.” *Psychological Bulletin* 112 (July 1992): 155–59.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. Daniel Riffe, “2008–2009 Annual Report” (paper, Association for Education in Journalism and Mass Communication annual meeting, Denver, August 2010).
2. Maria Elizabeth Grabe and Bruce H. Westley, “The Controlled Experiment,” in *Mass Communication Research and Theory*, ed. Guido H. Stempel, David H. Weaver, and G. Cleveland Wilhoit (Boston: Allyn & Bacon, 2003), 267–98.
3. W. James Potter, *On Media Violence* (Thousand Oaks, CA: Sage, 1999), 183.

4. Seth Geiger and John E. Newhagen, "Revealing the Black Box: Information Processing and Media Effects," *Journal of Communication* 43 (December 1993): 42–50.
5. Chen-Chao Tao and Erik P. Bucy, "Conceptualizing Media Stimuli in Experimental Research: Psychological versus Attribute-Based Definitions," *Human Communication Research* 33 (October 2007): 397–426; Daniel J. O'Keefe, "Message Properties, Mediating States, and Manipulation Checks: Claims, Evidence, and Data Analysis in Experimental Persuasive Message Effects Research," *Communication Theory* 13 (August 2003): 251–74.
6. S. Shyam Sundar, "Theorizing Interactivity's Effects," *Information Society* 20 (August 2004): 385–89.
7. Pamela J. Shoemaker, James William Tankard, and Dominic L. Lasorsa, *How to Build Social Science Theories* (Thousand Oaks, CA: Sage, 2003), 37.
8. Rebecca W. Hamilton and Debora Viana Thompson, "Is There a Substitute for Direct Experience? Comparing Consumers' Preferences after Direct and Indirect Product Experiences," *Journal of Consumer Research* 34 (December 2007): 546–55.
9. Rachel A. Smith, Timothy R. Levine, Kenneth A. Lachland, and Thomas A. Fediuk, "The High Cost of Complexity in Experimental Design and Data Analysis," *Human Communication Research* 28 (October 2002): 515–30.
10. W. James Potter, Roger Cooper, and Michel Dupagne, "The Three Paradigms of Mass Media Research in Mainstream Communication Journals," *Communication Theory* 3 (November 1993): 317–35; Glenn G. Sparks, "Is Media Research Prescientific? Comments Concerning the Claim That Mass Media Research Is 'Prescientific': A Response to Potter, Cooper, and Dupagne," *Communication Theory* 5 (August 1995): 273–80; W. James Potter, Roger Cooper, and Michel Dupagne, "Is Media Research 'Prescientific'? Reply to Sparks's Critique," *Communication Theory* 5 (August 1995): 280–86; Glenn G. Sparks, "Is Media Research Prescientific? A Final Reply to Potter, Cooper, and Dupagne," *Communication Theory* 5 (August 1995): 286–89.
11. Annie Lang, "The Logic of Using Inferential Statistics with Experimental Data from Nonprobability Samples: Inspired by Cooper, Dupagne, Potter, and Sparks," *Journal of Broadcasting & Electronic Media* 40 (summer 1996): 422–30.
12. Michael D. Basil, William J. Brown, and Mihai C. Bocarnea, "Differences in Univariate versus Multivariate Relationships: Findings from a Study of Diana, Princess of Wales," *Human Communication Research* 28 (October 2002): 501–14.
13. See, e.g., John B. Copas and H. G. Li, "Inference for Non-random Samples," *Journal of the Royal British Statistical Society: Series B (Statistical Methodology)* 59 (January 1997): 55–95; Eugene S. Edgington, "Statistical Inference and Non-random Samples," *Psychological Bulletin* 66 (December 1966): 485–87.
14. Julie L. Andsager and H. Allen White, *Self versus Others: Media, Messages, and the Third-Person Effect* (New York: Routledge, 2009).
15. Basil, Brown, and Bocarnea, "Differences in Univariate versus Multivariate Relationships," 501–14.
16. Carmen R. Wilson VanVoorhis and Betsy L. Morgan, "Understanding Power and Rules of Thumb for Determining Sample Sizes," *Tutorials in Quantitative Methods for Psychology* 3 (September 2007): 43–50.

17. Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner, "G*Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences," *Behavioral Research Methods* 39 (May 2007): 175–91.
18. Several journals require the reporting of observed power as well, which we think is an excellent idea. We also think it is useful to report effect sizes and observed power for nonsignificant findings. This way, the reader can assess whether the nonsignificant result is the result of low power.
19. Timothy R. Levine and Craig R. Hullett, "Eta Squared, Partial Eta Squared, and Misreporting of Effect Size in Communication Research," *Human Communication Research* 28 (October 2002): 612–25.
20. See, e.g., Geoffrey Keppel and Thomas D. Wickens, *Design and Analysis: A Researcher's Handbook* (Upper Saddle River, NJ: Pearson Prentice Hall, 2004); Stephen Olejnik and James Algina, "Generalized Eta and Omega Squared Statistics: Measures of Effects Size for Some Common Research Designs," *Psychological Methods* 8 (December 2003): 434–47.
21. We fully realize that not all mass communication experimental stimuli are "messages." However, we are specifically talking about variations in messages.
22. James J. Bradac, "Threats to Generalization in the Use of Elicited, Purloined, and Contrived Messages in Human Communication Research," *Communication Quarterly* 34 (winter 1986): 55–65; Grabe and Westley, "Controlled Experiment"; Sally Jackson and Scott Jacobs, "Generalizing about Messages: Suggestions for Design and Analysis of Experiments," *Human Communication Research* 9 (December 1983): 169–81; Sally Jackson, Daniel J. O'Keefe, Scott Jacobs, and Dale E. Brashers, "Messages as Replications: Toward a Message-Centered Design Strategy," *Communication Monographs* 56 (December 1989): 364–84; Byron Reeves and Seth Geiger, "Designing Experiments That Assess Psychological Responses," in *Measuring Psychological Responses to Media*, ed. Annie Lang (Hillsdale, NJ: Lawrence Erlbaum, 1994), 165–80.
23. Glenn Leshner, Paul Bolls, and Erika Thomas, "Scare 'Em or Disgust 'Em: The Effects of Graphic Health Promotion Messages," *Health Communication* 24 (July 2009): 447–58.
24. Andrea Miller, "Watching Viewers Watch TV: Processing Live, Breaking, and Emotional News in a Naturalistic Setting," *Journalism & Mass Communication Quarterly* 83 (autumn 2006): 511–29.
25. Renee Martin-Kratzer and Esther Thorson, "Use of Anonymous Sources Declines in U.S. Newspapers," *Newspaper Research Journal* 28 (spring 2007): 56–70.
26. O'Keefe, "Message Properties, Mediating States, and Manipulation Checks."
27. Henry L. Roediger III and Jeffrey L. Karpicke, "The Power of Testing Memory: Basic Research and Implications for Educational Practice," *Perspectives on Psychological Science* 1 (September 2006): 181–210.