Strategic and Tactical Limits of Comparison Studies

James M. Johnston
Auburn University

A comparison study is an experiment whose primary purpose is to compare directly (regardless of experimental design) at least two different procedures for changing behavior or two or more components of such a procedure. This paper argues that, in spite of their popularity, such studies typically lead to inappropriate inferences with poor generality based on improper evidence gathered in support of the wrong question, thus wasting the limited experimental resources. The discussion considers problems concerning the functions of comparison studies, the nature of the comparisons that are attempted, the generality of their findings, and the limited role that they can play in technological research.

Key words: comparison study, experimental question, generality, applied research, analytical study

For purposes of this discussion, a comparison study is defined as an experiment whose primary purpose, regardless of other methodological features, is to compare directly at least two different procedures for changing behavior or two or more components of such a procedure. Although it is true that all types of experiments necessarily involve comparisons at a tactical level, comparison studies (unlike analytically-oriented experiments) also have that focus at a strategic level. In other words, comparisons between whole procedures or between components of procedures are the central reason that comparison studies are conducted, whereas analytical studies are generally designed to identify and understand the nature of controlling variables, even though both types of experiments necessarily make comparisons between responding under treatment versus no-treatment conditions.

The typical experimental format for comparison studies involves running each procedure through its paces while measuring some sort of behavioral outcome, and this comparison focus can be pursued with either between-group or within-subject designs. (In fact, the type of experimental design used has little to do with the arguments in this paper, although between-groups designs do create special difficulties that will be noted.) If the arrangement of conditions to be com-

pared is sequential, the two procedures may sometimes use the same subjects. Otherwise, a different though presumably similar sample from the population of interest is usually used for each procedure. The ostensible reason for the comparison is to see which procedure (or procedural component) is better or more effective along some dimension for some usually very general set of applied circumstances, and the conclusion customarily attempts to announce a clear winner.

Examples of this experimental approach are easily found in any area of the literature devoted to developing and evaluating methods for changing behavior. For instance, Mosk and Bucher (1984) compared stimulus shaping to traditional prompting procedures for teaching visual-motor skills to retarded children, Barrera and Sulzer-Azaroff (1983) compared oral and total communication training programs with echolalic autistic children, and Repp, Barton, and Brulle (1983) compared two procedures for programming differential reinforcement on other behaviors.

The comparison study has always been a popular type of experimental strategy, especially in the applied literature. It appears to be a relatively easy type of experiment to design and conduct, it has intuitive appeal to non-professionals who may be interested in the results, and it is often selected when there is some need to evaluate a program's effectiveness. There is also probably a general tendency to think of experimentation in a com-

Reprints may be obtained from the author, Department of Psychology, Auburn University, Auburn, AL 36849.

parison framework because of the history with groups-comparison designs and inferential statistics that we in the social sciences all share, although, again, the important methodological issues that comparison studies raise are relevant for both between-groups and within-subject designs.

These tests are a professional version of the popular technological shootout, not unlike those conducted to compare consumer goods like toaster ovens, washing machines, or cars, and they have a face validity that makes them appealing to both consumers and investigators. However, while toaster ovens and behavior change methods are both technological products marketed to certain consumer groups, the appropriateness of comparison studies in their development, evaluation, and advertising is often quite different.

In fact, it is the thesis of this paper that comparison studies as discussed here are the bane of the applied literature. They often lead to 1) inappropriate inferences 2) with poor generality 3) based on improper evidence 4) gathered in support of the wrong question, 5) thus wasting the field's limited experimental resources. This is a strong indictment for so common an experimental format, but the problems with comparison studies are fundamental. They involve the nature of the experimental question, the nature of the comparison, and constraints on the generality (and thus the utility) of the conclusions. Furthermore, this list assumes that all other aspects of experimental method, such as measurement and experimental design, are properly treated so that they do not exacerbate these problems (in other words, that such studies are internally valid in the terms of Campbell & Stanley, 1963). The problems that plague comparison studies will be considered in the following sections.

Function of the Experiment

First, consider the difficulties surrounding the experimental question, the basic reason for doing the study. In the most fundamental sense, the guiding

question in comparison studies is often not really about behavior. While behavior seems to be a rather obvious focus, the question is usually about methods of controlling behavior. There is nothing wrong with this, of course; most applied research is about methods or procedures, and in some ultimate sense it must be. Nevertheless, there is a subtle but very significant difference in focus between applied studies that investigate behavior versus studies that investigate procedures for controlling behavior, even though both are aimed at the same result—effective technology. (Although all experiments “use” procedures, an important strategic distinction may be made between using them to study behavior or using them to study the procedures themselves.)

This point will be explored further below, but it is important here to appreciate that among all kinds of applied studies, comparison studies are especially procedural in nature. They simply ask which of two procedures is more effective in producing some kind of behavior change; there is usually little central interest in learning much else about behavior, although it is always possible as a bonus if the experimenter cares to analyze the data in that light.

There is another sense in which at least some comparison studies are not really asking about behavior, and this has to do with the routine details of experimental methods (measurement, data processing, and so forth). Behavior exists only between individual organisms and their environments, and in order to be effective, experimental methods must respect this biological fact (see Sidman, 1960, chapter 2). This means that if experimenters do anything that contaminates, dilutes, or otherwise distorts measures of behavior change, there is likely to be some deleterious effect on the inferences that can be drawn from the data. Among other actions, this caveat clearly includes the variety of measurement and data-processing techniques that result in collating individual data into some group amalgam.

Many comparison studies (virtually all

that use groups-comparison designs and even many that use within-subject designs) run afoul of this prohibition in one way or another. While the consequences do not necessarily hinder conclusions about behavior change *procedures*, they are quite threatening to inferences about *behavior*. In fact, based on the extent to which the fundamental nature of behavior is respected by experimental methods, an important distinction can be made between pure versus quasi-behavioral research (Johnston & Pennypacker, 1986). A large proportion of comparison studies are quasi-behavioral in nature (although analytical studies are hardly immune from these problems). That is, comparison studies look as if they are about behavior (after all, their data-base originated with observations of behaving individuals), but in the required methodological sense they often are not. Thus, the label "quasi-behavioral" seems appropriate.

It may be easier to view most comparison studies as actuarial in nature. Although the formal processes may deviate somewhat from those of the insurance industry, the general interest seems the same—to estimate dividends and risks for general categories based on statistical records alone, that is, without attempting to understand the reasons for each event so as to allow prediction in a more individualized fashion. This often seems the real reason why comparison studies are conducted—to give the user an experimental basis for predicting that one procedure will yield more or better changes in behavior than another for a typically broad class of individuals. When that experimental justification comes from studies that focus on merely measuring the degree or extent of effect on a sample rather than attempting to identify and understand the variables that make a procedure succeed or fail in each case, it is more an actuarial than a behavior-analytic study.

Even this clarification of motives may sometimes be generous. Too often, the real agenda may be political, in other words, advocacy of a prior conviction about the procedures being evaluated.

The researcher's interest not infrequently seems to be primarily aligned with only one of the procedures, and it may even be equally strongly in opposition to the competing procedure. The focus of the researcher's interest is usually unmistakably clear from his or her prose alone, but the selection and implementation of the competing procedure often removes any doubt. Not only is the comparison procedure sometimes chosen partly to make the procedure of primary interest look good (by choosing a "traditional" or a "no-treatment" procedure, for example), its implementation may not receive the same degree of attention as does the primary procedure. If this is not enough, it is characteristic of advocacy style research that various features of experimental method (such as measurement procedures) are knowingly or unknowingly bent in favor of the researcher's convictions, further compromising the veracity of any conclusions (see Johnston & Pennypacker, 1980, chapter 20).

Nature of the Comparison

While all of these problems are certainly worthy of concern, they can in principle be either tolerated or remedied in one way or another. Unfortunately, the remaining problems with comparison studies are more serious and bring into question the propriety of this type of experimental strategy. One such difficulty concerns how the nature of the two or more conditions chosen for comparison often raises the knotty questions of fairness and meaningfulness.

Because each procedure tested becomes the standard by which the other's effects are measured, selecting the proper procedures for comparison is critical. Certainly there must be a fair and meaningful basis for comparing two or more procedures. Usually the similarity of their intended function is the rationale; however, even this obvious selection criterion may be misleading. Although two procedures may address the same general behavioral goal, a number of detailed differences between them may often make each an inappropriate metric for the oth-

er. These differences may include the exact characteristics of the populations and settings with which each works best, the target behaviors and their controlling influences, or a variety of more administrative considerations such as the characteristics of the personnel conducting each procedure. If any of these or other critical features are not optimal for each procedure being compared, then it inevitably places each at an absolute disadvantage in producing its best effects, and it probably places one of the two at a relative disadvantage as well (a point also made by Van Houten, 1987).

For example, two programs designed to teach young children to read might not be meaningfully compared because they are unknowingly appropriate for children with different prerequisite skill levels. Contrasting them using the same population of subjects may imply that one is more effective than the other, although in fact each may work equally well when properly applied to slightly different subsets of the young, illiterate population. Although it might be argued that there can always be such differences that affect the results of compared procedures, studies with the comparison focus described here are less likely to discover such critical factors than are analytical studies aimed at identifying controlling variables, thus permitting erroneous conclusions to be made without qualification. In other words, if anything about the task or each's application makes one procedure more appropriate than the other, then any comparison is at least misleading or, in the worst case, entirely meaningless. For instance, you can compare a word processor to a typewriter for the task of writing a paper, but it is meaningless and unfair to compare a word processor to a calculator because they are not fully and equally appropriate for that purpose.

Understanding enough about two procedures to select them for a comparison study clearly requires each to have a considerable experimental history with which the prospective researcher is quite familiar. These literatures should identify through experimental analysis at least

most if not all of the *important* variables and any of their critical parameters required for each procedure to produce optimal effects. Any comparison conducted before such a literature is available is likely to misrepresent the procedure to some degree by not being able to implement it in its most effective form. The rigors of this obligation may contribute to the popularity of comparison conditions about which relatively little is known. These are the procedures usually labeled "traditional" and "no treatment," and they are almost inevitably a weak choice as a standard by which to evaluate some other procedure, however obvious the political benefits of the test.

A no-treatment condition is a blatantly inappropriate standard on at least two counts. First, such conditions do not represent the effects of a condition of nothingness—an environmental void. Whatever the no-treatment subjects are exposed to is simply not specified, usually because it is not known to the degree that the competing procedure can be described. This can hardly be a meaningful standard of comparison. Second, even if such a condition can be reasonably described, it is still a weak basis for comparison because it is likely to be the easiest possible criterion. All the procedure of primary interest has to do in order to be judged effective is to produce an appropriate disturbance of the status quo, hardly a crowning technological achievement. If at least this much has not been learned about a procedure through its experimental development, it seems pointless to attempt to certify its utility with such a bogus contest. Furthermore, the use of no-treatment comparison conditions tends to encourage concerns about the advocacy nature of the proceedings, there being less than equivalent procedural interest in the no-treatment condition.

Using some traditional procedure as the standard of comparison runs afoul of the same two difficulties to only a slightly lesser degree. Although the status of some condition as traditional may be unimpeachable, exactly what constitutes that procedure is usually unclear, partly be-

cause its popularity insures considerable procedural variety from case to case and partly because it usually lacks a serious, analytical, experimental history. Procedures that have been around long enough and please enough users to be called traditional may also be less effective than most prefer, which makes them a weak comparative standard and thus an easy way to demonstrate the relative effectiveness of some other procedure. Indeed, it is the general ineffectiveness of traditional procedures that often motivates the development of the technologies that they are then called on to assess.

A good example of such tactics is in the literature on individualized methods of college-level instruction. A sizeable portion of that literature is devoted to comparisons of various more or less individualized teaching procedures (such as the Personalized System of Instruction) with the traditional lecture method, clearly for the purpose of demonstrating or, more neutrally, assessing the relative effectiveness of the individualized methods. Given the evidence, conclusions have regularly been drawn about the superiority of individualized methods over the conventional lecture method (i.e., Hursch, 1976; Johnson and Ruskin, 1977; Lloyd, 1978). However, as any experienced college teacher knows, there is no such teaching "procedure" as the lecture method. Although most teachers do indeed talk in front of legions of adoring students a few times each week, it strains the denotation of the word lecture to apply it to all that goes on in these classes. More importantly, there is even less commonality in all of the other features of these "lecture method" courses, such as with critical variables like testing procedures and course contingencies. However, its popularity as a comparison condition for new teaching procedures is understandable because it is apparently easily beaten by other methods, a fact that may be contributed to by each researcher's freedom in using whatever arrangement of the "lecture procedure" seems to be appropriate in each case. Any value in this victory seems more political than technological (Johnston, 1975).

In summary, the requirement that the two or more procedures in question be meaningfully and fairly comparable is not easy to satisfy. The only way to address it properly leads experimentation away from comparison studies and in the direction of analyses of sources of control over a procedure's effects. Understanding what makes a procedure more or less effective requires a program of experimental analysis that identifies 1) those features of the target behaviors and their histories that can interact with procedural components in some way, 2) those features of the circumstances under which the procedure is applied that can modulate its behavioral effects, and 3) those procedural options and their parameters that can influence behavioral outcomes. This considerable requirement should suggest that comparison studies (and, for that matter, all evaluation studies) are usually premature, at best. The proper question is not "Which is better?" but "What are the variables that each requires to produce optimal effects?"

Generality

Aside from the appropriateness of the comparison, there is another major problem that depreciates the value of conclusions from comparison studies, and it concerns their generality or, in the terms of Campbell and Stanley (1963) and Cook and Campbell (1979), external validity.¹ In spite of the fact that the motivation underlying comparison studies is primarily to reach a conclusion that one of the procedures is "better" across the broadest of circumstances, the generality of most comparison studies must be evaluated as poor. The reasons are especially obvious when traditional and no-treatment conditions are used as comparisons, but they plague most behavioral procedures that show up in comparison studies, even those with impressive research histories.

In order for conclusions about the relative effectiveness of two procedures to

¹ Portions of this argument were included in Johnston (1985).

have good generality to other applications of each procedure, the critical features of each (including procedural elements, parameters of important variables, required characteristics of target behavior and clients, and the necessary degree of control over setting features) must be known and then held constant across all routine applications. Because the first requirement is rarely accomplished, the second is rarely possible. In other words, unless these critical variables that actually make a procedure work have already been identified and publicized in the existing literature, they cannot be properly and consistently managed across its many applications. These admittedly challenging requirements make any effort to establish one procedure as "better" than another (or, worse yet, "best" with regard to some very broad range of circumstances) an improper and unattainable goal in the absence of such a literature. Furthermore, whether we find such requirements reasonable or attainable for the present state of the field has no bearing on their importance, which stems from the nature of behavior.

Even if one tried to implement a procedure in the same way across repeated applications, it would be difficult to avoid accidentally changing some important variables. However, in practice each time a procedure is applied, numerous variations in its components and their circumstances are unknowingly made, willingly accepted, and intentionally incorporated (see Barber, 1976, chapters 3 and 7, for a discussion of this problem in research contexts). When this reality is combined with the substantial lack of an adequate experimental rationale for these changes or any evidence in the literature about their different effects, there may be little basis for even describing all such applications with the same label. The result is the emergence of very large, amorphous, and overlapping categories of procedures. The use of simple procedural labels is clearly gratuitous and implies a uniformity that is quite misleading.

For instance, applications labeled "Personalized System of Instruction" collectively include an impressive vari-

ety of components, although each instance often does not even incorporate all of the five defining features specified by Keller (1968). This same variety may be seen for almost all of the field's most common procedures, ranging from the simple (such as differential reinforcement of other behavior) to the complex (such as token economies). Many such labels have become rather crude categorical references with no more than the general conversational value of other abstractions.

Of course, there is nothing wrong with variations on a particular procedural theme; in fact, variations that have been examined by analytical research programs and that allow the practitioner to adjust a procedure to accommodate unique circumstances in the field without damaging its effectiveness are extremely valuable. For instance, an analytical program of research concerning rumination in retarded individuals shows that feeding satiation quantities of food at meals will greatly decrease ruminating, although it is clear that this is an inappropriate technique when the resident is at or above his or her normal body weight. However, this research also shows how satiation feeding can be modified or dropped entirely while achieving the same effect by incorporating a number of procedural variations (Rast, Johnston, & Drum, 1984; Rast, Johnston, Drum, & Conrin, 1981; Rast, Johnston, Allen, & Drum, 1985; Rast & Johnston, *in press*).

Unfortunately, the lack of such experimental analyses for most procedural applications and their labels means that there is insufficient evidence that each application is effective by virtue of specified critical features. The procedural variety instead suggests that any effectiveness is due to unique combinations of at least partly unknown influences that probably go beyond those uniformly recognized as defining the procedural label.

Poor generality of comparison studies is likely, then, for two reasons. Each of the procedures being compared is a naively unique arrangement of variables, and every subsequent application is also unique with regard to its own arrange-

ment of functional variables. Generality of method accrues not to labels but only to the particular combination of variables that makes each procedure effective and that will guarantee its repeated effectiveness as often as those elements are brought together. To the extent that those elements are unknown or only suspected, each application risks violating the conditions that define the version used in the comparison study, thus constraining the correctness of the prediction suggested by the comparison study. Generality of method is effectively gained only by discovering and understanding the role of critical variables whose status in particular applications can then be used to predict the effects of the version of the procedure being planned (Johnston & Pennypacker, 1980, chapter 19).

In the terms of Cook and Campbell (1979), the differences between the features of the procedures being compared and the different versions of them subsequently used in other studies or in clinical settings are the source of threats to the external validity of such studies (although the extensions by Judd & Kenny, 1981, of the concept of construct validity originated by Cronbach & Meehl, 1955, should also be noted). However, the present points about generality differ from the usual discussions of external validity (see, for instance, Kazdin, 1982). The argument here is that there is no way to sufficiently specify subject, setting, and procedural characteristics to insure reliable generality of comparison study conclusions because the critical variables responsible for the effects of each procedure are unlikely to be thoroughly identified in this style of research (in part, because it is not designed to do so). This knowledge is therefore not available to guide or limit either statements of generality or efforts to engineer generality by carefully arranging those features of any application of a study's findings.

Limited Role of Comparison Studies

The indictment has thus far characterized comparison studies as generally being not fundamentally behavioral, not mean-

ingfully comparative, not fairly conducted, not productively evaluative, not very general in their findings, and not very high on the list of research priorities. Perhaps the only remaining question is, "Is there any hope for them?"

The answer is a qualified yes, but first it is important to distinguish comparison from evaluation. If there is a need to evaluate or assess the effects of a procedure, this need in no way requires or encourages comparison with another procedure. Evaluation only asks about the effects of a procedure under certain field conditions, and simple description is all that is called for. Whether a program "works" can be measured against standards specified by carefully defined goals that the target behavior, setting, and cultural context supply. This is the only comparison that is meaningful if the purpose of the effort is simply to evaluate the applied effectiveness of a newly developed or differently applied procedure.

What, then, is left for comparison studies to do? To compare, obviously, but only after acknowledging that such studies are not really for asking about behavior or about evaluation. Instead, they are primarily appropriate for answering questions about the relative effects of two or more procedures for a particular combination of target behavior, client, and setting variables for relatively narrow or parochial purposes. Furthermore, although comparison studies can certainly be conducted for only one or a very small number of subjects, more often they are justified by the need to make a procedural selection for a larger population, such as all children in a school system. However, the motivating interests should be more practical or even logistical than generic; concern should focus on which of two alternatives is the more applicable and effective choice rather than which is the best in some overall or general sense.

As already argued, answering the question about which of some group of procedures is generally superlative is usually impossible because it is frequently not really a meaningful question. In addition, there may be fully as much interest in the requirements of the procedures from an

administrative perspective as in their relative behavioral effectiveness (see Hopkins, 1987). For instance, procedure A may be somewhat more effective, while procedure B may require fewer personnel or less skilled personnel or less money or easier scheduling or any of a number of other practical advantages. These considerations are often no less important than the procedures' behavioral effects because the best choice in terms of behavioral effectiveness may be less effective than its competitor if its administrative requirements cannot be met when it is routinely applied.

Because of this parochial focus that eschews interest in conclusions with general applicability, comparison studies as approached here are unlikely to be conducted for the purpose of being published as statements of the generality of a procedure, and the availability of such data is not likely to encourage a post hoc decision to write a paper that could offer only conclusions of intentionally narrow and local import. This may be why most of the literature that makes comparisons generally does so in the traditional style critiqued in this paper. As a result, actual examples of the alternative style proposed here are not readily available in the archival literature.

Fully as important as understanding the proper role of comparison studies is understanding their priority or place in a procedure's experimental history. They should be the very last type of investigation that a procedure encounters. They are appropriate only after the proper course of experimental development, analysis, refinement, and evaluation has been completed, and this means an experimental history that is more thematic and analytically sophisticated than is customarily the case in today's applied literature. Meeting this challenge requires a rather lengthy and integrated series of experiments that analyze behavior more than procedures by focusing on identifying the variables that do or can control behavior and examining the ways in which those variables can most effectively be arranged in practical form. With such a history, the constraints of com-

parison studies are not disappointing because all of the other questions (concerning the identification of critical procedural, subject, and setting variables, the administrative requirements, and the nature of behavioral effects) have already been answered.

With this kind of experimental lineage, the question of which of two procedures is better for a particular type of behavioral problem under a particular set of circumstances simply should not come up very often. The reason is that when enough is known about the critical variables required for each procedure's maximally effective application, it will usually be clear that one procedure is more appropriate than another for the particular combination of features that characterize the behavioral problem and its attendant circumstances. When this is the case, there is no meaningful and fair comparison that can be or will need to be made.

Unfortunately, such a literature is not typically available in any of the many areas in which we are now called upon to provide behavior change services, and the final defense of comparison studies is that they are at least needed to guide programmatic decision-making, however crudely, until this more sophisticated database is available. It is true, of course, that even with their limitations comparison studies are better than no studies at all. However, this argument runs into another serious problem in the applied field—the terribly small proportion of its participants who are interested in conducting or employed to conduct any kind of research at all. Although no figures are available, it would seem that only a precious few of those earning graduate degrees from behavior analysis programs are really very interested in research careers, and only some fraction of them are interested in an experimental focus on applied problems. Of these, only a small minority wind up being employed under circumstances in which research is even possible or encouraged, and only some of these circumstances are conducive to conducting sound behavioral research, whatever its style. To propose de-

voting any portion of this tragically small pool of scientific resources to continuing an experimental tradition characterized more by its limitations than by its strengths seems shortsighted, especially when the experimental alternative is so rich with immediate potential for improving our technological abilities.

REFERENCES

- Barber, T. X. (1976). *Pitfalls in human research*. Elmsford, NY: Pergamon Press.
- Barrera, R. D., & Sulzer-Azaroff, B. (1983). An alternating treatment comparison of oral and total communication training programs with echolalic autistic children. *Journal of Applied Behavior Analysis*, 16, 379-394.
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand-McNally.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Cook, T. D., & Campbell, D. T. (Eds.) (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand-McNally.
- Hopkins, B. L. (1987). Comments on the future of Applied Behavior Analysis. *Journal of Applied Behavior Analysis*, 20, 339-346.
- Hirsch, D. E. (1976). Personalized systems of instruction: What do the data indicate? *Journal of Personalized Instruction*, 1(2), 91-105.
- Johnson, K. R., & Ruskin, R. S. (1977). *Behavioral instruction: An evaluative review*. Washington, D.C.: American Psychological Association.
- Johnston, J. M. (1975). Alphabet soup, generality, and the good old days. In James M. Johnston (Ed.) *Research and technology in college and university teaching* (pp. 9-13). Champaign, IL: Research Press (distributor).
- Johnston, J. M. (1985). Controlling professional behavior: A review of *The effects of punishment on human behavior* by Axelrod and Apsche. *The Behavior Analyst*, 8, 111-119.
- Johnston, J. M., & Pennypacker, H. S. (1980). *Strategies and tactics of human behavioral research*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Johnston, J. M., & Pennypacker, H. S. (1986). Pure- versus quasi-behavioral research. In A. Poling, R. W. Fuqua, & R. Ulrich (Eds.) *Research methods in applied behavior analysis: Issues and advances* (pp. 29-54). New York: Plenum.
- Judd, C. M., & Kenny, D. A. (1981). *Estimating the effects of social interventions*. New York: Cambridge University Press.
- Kazdin, A. E. (1982). *Single-case research designs*. New York: Oxford University Press.
- Keller, F. S. (1968). "Goodbye teacher . . ." *Journal of Applied Behavior Analysis*, 1, 79-89.
- Lloyd, K. E. (1978). Behavior analysis and technology in higher education. In T. A. Brigham & A. C. Catania (Eds.) *Applied analysis of social and educational behavior* (pp. 482-521). New York: Irvington Publishers.
- Mosk, M. D., & Bucher, B. (1984). Prompting and stimulus shaping procedures for teaching visual-motor skills to retarded children. *Journal of Applied Behavior Analysis*, 17, 23-34.
- Rast, J., & Johnston, J. M. (In press). Effects of pre-meal chewing on ruminative behavior. *American Journal of Mental Retardation*.
- Rast, J., Johnston, J. M., Allen, J., & Drum, C. (1985). Effects of nutritional and mechanical properties of food on ruminative behavior. *Journal of the Experimental Analysis of Behavior*, 44, 195-206.
- Rast, J., Johnston, J. M., & Drum, C. (1984). A parametric analysis of the relationship between food quantity and rumination. *Journal of the Experimental Analysis of Behavior*, 41, 125-134.
- Rast, J., Johnston, J. M., Drum, C., & Conrin, J. (1981). The relation of food quantity to rumination behavior. *Journal of Applied Behavior Analysis*, 14, 121-130.
- Repp, A. C., Barton, L. E., & Brulle, A. R. (1983). A comparison of two procedures for programming the differential reinforcement of other behaviors. *Journal of Applied Behavior Analysis*, 16, 435-445.
- Sidman, M. (1960). *Tactics of scientific research*. New York: Basic Books.
- Van Houten, R. (1987). Comparing treatment techniques: A cautionary note. *Journal of Applied Behavior Analysis*, 20, 109-110.