JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1982, 15, 477-492 NUMBER 4 (WINTER 1982)

THE INTEGRITY OF INDEPENDENT VARIABLES IN BEHAVIOR ANALYSIS

LIZETTE PETERSON, ANDREW L. HOMER, AND STEPHEN A. WONDERLICH

UNIVERSITY OF MISSOURI-COLUMBIA AND
MISSOURI DEPARTMENT OF MENTAL HEALTH

Establishing a functional relationship between the independent and the dependent variable is the primary focus of applied behavior analysis. Accurate and reliable description and observation of both the independent and dependent variables are necessary to achieve this goal. Although considerable attention has been focused on ensuring the integrity of the dependent variable in the operant literature, similar effort has not been directed at ensuring the integrity of the independent variable. Inaccurate descriptions of the application of the independent variable may threaten the reliability and validity of operant research data. A survey of articles in the *Journal of Applied Behavior Analysis* demonstrated that the majority of articles published do not use any assessment of the actual occurrence of the independent variable and a sizable minority do not provide operational definitions of the independent variable. The feasibility and utility of ensuring the integrity of the independent variable is described.

DESCRIPTORS: reliability, independent variable manipulation, methodology, implementation reliability, validity

The primary goal of behavior analysis has been described as going beyond the simple demonstration of changes in behavior to include the demonstration that changes in the target behavior are functionally related to changes in the environment (Baer, Wolf, & Risley, 1968). This typically involves ensuring that changes in the dependent variables are due to systematic changes in the independent variable, rather than to changes in any uncontrolled extraneous variables (Sidman, 1960). Precise, demonstrated control of the independent variable and measurement of the dependent variables is therefore necessary for demonstrating the existence of a functional relationship.

In applied behavior analysis, rigorous and calibrated electromechanical recording of the dependent variable is frequently not as possible as in an animal laboratory. Thus, one of the earliest demands on applied researchers in behavior analysis was for accurate and reliable description and observation of the dependent variable (e.g., Heyns & Lippitt, 1954). The classic literature of observational technology (e.g., Arrington, 1943; Bijou, Peterson, & Ault, 1968; Rosenthal & Rosnow, 1969; Webb, Campbell, Schwartz, & Sechrist, 1966; Weick, 1968) was joined in later years by a new literature, concerned with problems of dependent variable observation such as observer reactivity (e.g., Hanley, 1970; Romanczyk, Kent, Diament, & O'Leary, 1973), observer bias (e.g., McNamara & MacDonough, 1972; O'Leary, Kent, & Kanowitz, 1975), observer drift (e.g., O'Leary & Kent, 1973; Reid, 1970), observational complexity (e.g., Kazdin, 1977*a*) and

This research was supported by a Summer Research Fellowship granted to the first author by the University of Missouri-Columbia Research Council. The authors are grateful to Robert J. DeRubeis, Donald P. Hartmann, Steven D. Hollon, and Steven Richards for their help in locating research examples for this paper. Appreciation is extended to Samuel M. Deitz and James Johnston for their helpful comments on past versions of this manuscript. Reprints may be obtained from Lizette Peterson, Psychology Department, 210 McAlester Hall, University of Missouri-Columbia, Columbia, Missouri 65211.

various sources of inflation of observer reliability estimates (e.g., Hartmann, 1977). The number of articles in the *Journal of Applied Behavior Analysis* (*JABA*) that contain reliability estimates of the dependent variable have steadily increased in recent years (Hayes, Rincover, & Solnick, 1980). Indeed, it is unlikely that a paper without necessary dependent variable reliability estimates would be accepted for publication in *JABA*.

In applied behavior analysis, control of the independent variable is also more difficult than in an animal laboratory. It thus may be surprising to note that the methodological rigor applied to the observation of the dependent variable may not have been applied to the independent variable. Since independent variables frequently involve human judgments, some assessment of the reliability and accuracy of the judgments may be necessary. For example, Johnston and Pennypacker (1980) stated that:

> The independent variable must be represented by some environmental event, the physical parameters of which are known, specified, and controlled to the extent required. Such a clear description of the independent variable is essential if any factually accurate statement is to issue from the experimental effort (p. 39).

Inadequate assessment of the independent variable may thus render conclusions about the functional relationship between the dependent target behavior and the independent treatment variable suspect. For example, recent reviews of experimental results from token economies (Nelson & Cone, 1979) and Differential Reinforcement of Other Behavior procedures (DRO; Homer & Peterson, 1980) complained that the independent variables are not described in sufficient detail for evaluation of the results. These and other reviews (e.g., Hobbs, Moguin, Tyroler, & Lahey, 1980) have implicated the lack of independent variable description and verification as a likely cause for poor replication of results.

The methodological gap between the observation and reporting of the dependent variable in comparison to the independent variable may have serious consequences. A variety of questions need to be answered before any definitive conclusions on this issue can be reached. First, is there anything in the literature to suggest that accurate description and observation of the independent variable is important? Second, if accurate description and observation fail to take place, does this ever result in differences between the independent variable as presented in the research report method section and the independent variable as actually applied to the subject? Third, if such differences are found, is there any significant cost to the research findings and to the field? Finally, if a cost exists, is there a cost-effective way of preventing the problem in future research? This paper proposes some answers to these questions.

Methodological Statements on Independent Variable Accuracy

First, it has already been suggested that the operant literature clearly specifies the importance of establishing a functional relationship between the dependent variable and the independent variable by *observing* that the systematic manipulation of the latter results in changes in the former (e.g., Hersen & Barlow, 1976; Johnston & Pennypacker, 1980; Sidman, 1960). Observation of the dependent variable alone will allow unambiguous conclusions about changes in the target behavior, but it will not allow conclusions about the source of those changes (Billingsley, White, & Munson, 1980).

A curious double standard has developed in operant technology whereby certain variables (e.g., social behavior, smiling, and attention) routinely have operational definitions and some measure of observer reliability when the observed behavior is the target response or dependent variable (Milby, 1970; Reisinger, 1972; and Kazdin, 1973, respectively), but no such rigor is applied to the same behaviors when they appear as antecedents or consequences to the target behavior, as independent variables (Strain, Shores, & Timm, 1977; Dorsey, Iwata, Ong, & McSween, 1980; and Hasazi & Hasazi, 1972, respectively). Either such precautions as definition and reliability of observation are necessary or they are not. The observational literature clearly suggests that they are. It is unlikely that most readers would accept the experimenter's claim that "social responding increased when the treatment variable was applied," if no data on social responding were presented to substantiate the claim. How, then, is it possible to accept the experimenter's suggestion that "social responding by the treatment agent was increased and changes in the dependent variable occurred" with no further definition or observation of social responding? The simple statement that treatment was applied as outlined in the method section or treatment manual seems insufficient (DeRubeis & Hollon, 1981).

Furthermore, although observational technology has centered on the dependent variable, all statements made concerning the observation of the dependent variable are also applicable to the observation of the independent variable. Experimenter bias (e.g., McNamara & MacDonough, 1972) may, in fact, be more likely in observing the treatment variable than the dependent variable, since the treatment variable is predefined and an informed observer might simply report observing what the therapeutic agent was supposed to do rather than what was actually done. Similarly, if the therapy agent could discriminate when observation was taking place, it is likely that the agent would adhere more closely to the assigned treatment during observational periods, demonstrating reactivity (e.g., Romanczyk et al., 1973). Finally, Hersen (1981) noted that changes can and often do develop between different treatment agents and within the same agent across time. Treatment during the first phases of a study may only approximate treatment at a later date; if observers fail to recognize such a process, the result might be both treatment and observer drift (Reid, 1970). Thus, the behavior analytic literature clearly supports the importance of independent variable definition and assessment.

Other literatures have demonstrated even greater support for independent variable definition and assessment. Stallings (1975), for example, described the effects of both open and behavioral classrooms on children's learning. Stallings carefully measured both children's in-class behavior (dependent variables) and the teachers' administration of each treatment (independent variables). Although the consistency with which the treatments were applied was generally judged to be adequate, the consistency of actual treatment variable application compared to scheduled treatment variable application *within* any one classroom ranged from $r = .30$ to .96. There was even less consistency across different classrooms. Resnick and Leinhard, when critiquing Stallings' study (cited in Stallings, 1975), noted that many of Stallings' findings would be uninterpretable without the examination of the actual (not the planned) application of the independent variable. There are many other related research areas in which investigators have urged the definition and assessment of the independent variable, including a review of interventions on children's problem solving (Urbain & Kendall, 1980) and children's psychotherapy (Hartmann, Roper, & Gelfand, 1977), behavior therapy with psychotic adults (Paul & Lentz, 1977), pharmacotherapy (Becker & Schuckit, 1978) and psychotherapy with depressed adults (Rounsaville, Weissman, & Prusoff, 1981). There have also been repeated suggestions to document the application of the independent variable within the field of program evaluation (e.g., Cook & Campbell, 1979; Donabedian, 1966; Way, Lund, & Attkisson, 1978).

Reports and Suggestions of Independent Variable Inaccuracies

There is some evidence within applied behavior analysis to suggest that the treatment described in the method section may, in fact, differ from the actual applications of the treatment. This statement of the problem should not imply that the experimenter does not retain complete flexibility to alter the treatment strategy at will. The application of the treatment variable always remains at the discretion of the experimenter. The problem occurs when the experimenter believes that the application of the independent variable has certain physical and temporal properties when, in fact, different physical and temporal properties of the independent variable apply.

Six classes of results of independent variable inaccuracies will be reviewed briefly, including: (a) cases where a difference between programmed and actual independent variable application was noted during the study, prior to the formation of any conclusions, (b) cases in which inaccuracies were noted at the end of a study, but basic conclusions remain the same, (c) cases in which inaccurate administration of the independent variable changed the conclusion, causing an ineffective treatment to appear effective, (d) cases where independent variable inaccuracies rendered an effective treatment ineffective, (e) cases where failure to replicate was caused by faulty independent variable application, and (f) cases where failures to replicate are linked to independent variable implementation inaccuracies.

First, researchers may note independent variable inaccuracies during the course of the treatment phase. A treatment agent may either become lax with timing or with effortful treatment techniques or may begin to add techniques not prescribed by the experimenter to the treatment regimen. Any case in which the treatment agent gradually alters the treatment can be termed "therapist drift." This phenomenon has been observed, recorded, and reported by behavioral investigators (e.g., Bellack, Hersen, & Himmelhock, 1980; Hollon, Mandell, Bemis, DeRubeis, Emerson, Evans, & Kriss, Note 1). Many times therapist drift results in the treatment agent's increasing use of therapeutic techniques, as opposed to becoming progressively more lax with treatment variable application (DeRubeis, Note 2).

Not all inaccuracies in independent variable application are noted prior to a study's completion, however. For example, many studies in *JABA* have overtly analyzed attempts to alter the behavior of a treatment agent, who will, in turn, alter a target subject's behavior. Thus, some reports have routinely obtained data on both the independent variable and the dependent variable, and have noted some level of inaccuracy in independent variable application which is not corrected. In one such report, teachers were to present Distar materials at either a rapid or slow rate, and the main purpose of the study was to analyze the effects of rate of presentation. Teachers received clear mechanical cues for their rate of presentation. However, observation and recording of the independent variable of Distar presentation rate demonstrated that there was a great deal of variation in the actual rate of teacher presentation; in one case the rate during a "rapid presentation" was actually slower than in the "slow presentation" sequence (Carnine, 1976). This is a particularly cogent example because many studies in applied behavior analysis use teachers or other lay therapy agents *without* mechanical cues for independent variable application and these studies typically assume that the therapy agent is accurate. Other examples from applied behavior analysis include significant inaccuracies in teacher attention (Friedling & O'Leary, 1979), proctor instruction of trainees (Mathews & Fawcett, 1977), and parent application of behavioral techniques (e.g., Porterfield, Herbert-Jackson, & Risley, 1976, found a range of 33 to 100% actual adherence to the description of the independent variable). In fact, a sizable number of the studies in *JABA* which were reviewed and found to assess independent variable application noted differences in the planned administration of the independent variable and the actual application. Again, the issue is not that the application of the independent variable must be stable but rather that when it is not stable, the method section should not describe the application as rigid and consistent. Despite the inaccurate application of treatment in the studies just cited, beneficial effects of treatment were noted (although it is unclear whether the effects were more or less beneficial than would have resulted if the experimenter's specified method had been followed). In other cases, inaccuracy in independent variable application has resulted in more serious changes in treatment variable outcome.

Baer et al. (1968) described the case in which the treatment agent allows the subject a short time period in which to respond correctly during baseline and then (through accident or intention) allows more time for responding during the experimental phase. Baer et al. noted that even with an ineffective treatment, these failures to adhere to the uniform procedures as described in the method section might result in the appearance of treatment success. Since the experimenter would be likely to be satisfied with the results, there would be little reason to question the treatment agent's adherence to the application of the independent variable. Thus, a worthless treatment would be judged to be effective and only failures to replicate the results would correct the error. Since failures to replicate are published less often than successful intervention, such errors are likely to be corrected only slowly and at great cost (Homer & Peterson, 1980), with the cost sometimes being the rejection of behavior therapy techniques by community treatment agents (Hersen & Barlow, 1976).

Other failures to adhere to the experimenter's description of independent variable implementation may harm rather than enhance treatment effectiveness. This kind of alteration may be more likely to be noted by the experimenter than is a favorable alteration, but at times the intervention is completed before the reason for the lack of treatment effectiveness is clear. Bernal, Klinnert, and Schultz (1980), for example, found that although the parent-therapists receiving behavioral training reported more child improvement than parents receiving client-centered counseling, there were no differences in observations of actual child behavior. Further assessment of the behavior of the parent-therapists demonstrated that the parents trained in behavioral techniques were not applying those techniques any more than were the counseled parents. Such findings are not limited to lay treatment agents. For example, Wodarski, Feldman, and Pedi (1974) found no effects from a behavior modification treatment program. Fortunately, the investigators had obtained data on both the professional treatment agents' behavior and the children's behavior. Analysis of these data demonstrated that the professional treatment agents had failed to apply the specified behavioral treatments (contingent time-outs, praise, directions).

In some cases of clear failure to replicate certain experimental results, lack of accurate treatment implementation has been specifically implicated as a cause. For example, Fleishman (1981) described many attempts to replicate Patterson's (1974) behavioral treatment of noxious child behaviors (yelling, aggression). He noted that the differences in results from studies which successfully replicate or which fail to replicate Patterson's results may have been due to differences in the degree to which the independent variable was implemented as described. Since treatment variable implementation had not been typically observed and reported, however, the extent to which inaccurate implementation was responsible for differences is not clear. Fleischman ended his review by noting "In any case, it suggests the need for further research on not merely the impact of social learning derived technology per se on aggressive children but also on how that technology is applied" (pp. 350-351). Similarly, other investigators have noted explicitly that their failure to replicate may have been due to inaccuracies in treatment variable application (e.g., Bornstein & Quevillon, 1976).

Cost of Inaccuracy

Despite the apparently heavy cost of independent variable inaccuracy, some operant researchers may disavow any generality in these results and will continue to argue that, in general, there is not substantial cost to independent variable inaccuracy. Several arguments might be advanced here, including the argument that "the use of steady state responding guarantees the accuracy of treatment application," "there is always some amount of 'play' in the administration of the treatment variable and that only makes the functional relationship between dependent and independent variable more robust" and " if there was substantial inaccuracy in the application of the independent variable, the experimenter would immediately be informed of it through relevant changes in the dependent variable." Each of these typical arguments against a general cost of independent variable inaccuracies must be dismissed before the generality of these results can be assumed.

First, there is the claim that in general, steady states and behavior stability of the dependent variable would provide ample demonstration of a functional relationship between the independent and dependent variable, and that observation of the dependent variable alone is sufficient for such demonstration. The argument would go like this:

1. If the stable independent variable is functionally related to the dependent variable, steady-state responding in the dependent variable will occur.

2. Steady-state responding in the dependent variable occurs.

3. Therefore, the independent variable is functionally related to the dependent variable and is responsible for the steady state.

4. Therefore, the application of the independent variable must have been stable.

This line of reasoning has been labeled the Fallacy of Affirming the Consequent (Johnston & Pennypacker, 1980). Although this logic is weak, it can be used to demonstrate behavioral control if applied repeatedly and if complete assessment is made of both the independent and dependent variables in stable states and in transition to verify all four stages of logic empirically. However, if the experimenter cannot demonstrate the state of *both* independent and dependent variables, no conclusion can be reached.

Reflection may demonstrate that there are many cases of stable states in the dependent variable that are due to carefully programmed fluctuations in the independent variable. Drug tolerance is one common example; steadily increasing dosages of some drugs (independent variables) may be required to maintain a stable drug response (dependent variables). Similarly, the animal operant literature contains several examples of complicated schedules of fluctuating contingencies designed to produce steady-state responding (e.g., Dallemagne & Richelle, 1970). Thus, a stable dependent variable does not necessarily ensure a stable independent variable.

Second, it might also be argued that the experimenter never has complete command over all the controlling variables, since there is always likely to be some degree of background noise and extraneous variables present in addition to the independent variable. However, when changes in the independent variable (even if imperfectly applied) result in changes in the dependent variable, one might argue that this demonstrates a robust functional relationship and adds strength to the conclusion of experimental control. This argument can only be made if the dependent variable changes appropriately; the imperfect application of the independent variable may result in loss of control and this would scarcely suggest a robust functional relationship. Some such failures (which are rarely published and thus shared openly with other researchers) have been documented earlier. Even if changes in the dependent variable are obtained, this does not necessarily indicate a strong functional relationship between dependent and independent variables unless the experimenter can show that the changes were due to the treatment variables and not to other variables. In other words, any "extraneous" (nonindependent) variables must be demonstrated either to be random or to have effects in the opposite direction of the effects of the independent variable. However, many of the "extraneous" variables may not be random or "countertherapeutic" at all. Indeed, as has been documented earlier, "therapist drift" to include nontreatment variables is often the result of attempts to impact the target behavior in the same manner as the independent variable attempts to impact behavior. Such extraneous variables may ride "piggyback" on the independent variable application as, for example, a treatment agent who gradually lengthens a DRO interval (described in the method section as the independent variable) while maintaining the target subject's adaptive behavior with smiles and praise (extraneous variables not described in the method section). The "extraneous variables" may even replace the ostensible independent variable as could occur with a treatment agent who used self-attributing praise to a subject (an "extraneous" variable) in combination with apparently powerful but actually inert tokens (the independent variable of record). Thus, anticipated changes in behavior when the treatment variable of record is applied do not necessarily indicate a robust functional relationship.

The example just cited also explains how the third common argument against the potential cost of independent variable inaccuracy can be shown to be false. It is often concluded that if something went "wrong" with the treatment variable application, the experimenter would immediately be aware of this because of unforeseen changes in the dependent variable. When the extraneous variables covary with the independent variable, the dependent variable may reveal only the "expected" change. It will yield no information about the cause of that change. Thus, if the treatment agent inadvertently adds to the magnitude or frequency of the independent variable or uses adjunct techniques not described in the method section, changes in the dependent variable will suggest that the treatment is more effective than it actually is. Because the experimenter will be satisfied with the obtained results, no real assessment of the true functional relationship will take place. In contrast, the experimenter sometimes becomes aware of problems in experimental methodology because the dependent variable does not change as planned. However, because this lack of effect can be due to a number of factors, including subjects' past behavioral history, reinforcer effectiveness, schedule parameters, or extraneous variables which are truly random, the experimenter may not be able to determine whether or not the lack of effect was due to insufficient administration of the independent variable. Even a powerful treatment may thus appear impotent. Such failures, especially when they are not well investigated and documented, can have very detrimental effects in the applied community where disappointment with a procedure may lead to the tendency to eliminate the procedure entirely from therapeutic programs (Hersen & Barlow, 1976).

The possibility of accepting a powerless program as strong or rejecting a powerful program as having no effect is the ultimate cost of lack of independent variable accuracy. However, it might be argued that occasionally accepting an incorrect finding may simply be the price of living in an imperfect world; the inaccuracy of the finding will be brought to light later, as a failure to replicate. This argument is far too pat and it ignores what Hersen and Barlow (1976) term the "often discouraging and sometimes painful process of clinical trial and error" (p. 355) involved in attempts to replicate. If advancement of the science is made only through direct and systematic replication (Sidman, 1960), then allowing some number of incorrect reports in the literature with no information concerning the degree to which they may be inaccurate is extremely costly.

All that can be demonstrated *in general* is a potential cost (although an extravagant one), since with the present system there is no way of gauging the level of inaccuracy or the actual cost that may exist in the current literature. However, it is possible to determine the number of studies at risk for independent variable inaccuracy. Before concluding the discussion on cost and going on to the suggestion for possible solutions, however, it would seem appropriate to assess the degree to which investigations in applied behavior analysis are at risk for inaccurate treatment variable application. If the number of at risk studies is high, perhaps that finding will influence decisions on potential cost and solutions to that cost.

METHOD

Two independent observers were responsible for rating the articles in *JABA* from 1968 (Volume 1, Number 1) to 1980 (Volume 13, Number 3). The primary observer rated every issue and the reliability observer rated one issue (25%) per year. Only experimental articles were rated; "experimental" was arbitrarily defined as any article longer than three pages of text that included a method section. This definition excluded brief reports which might have gathered independent variable data but failed to report it because of the condensed nature of the presentation, as well as technical notes and theoretical presentations.

Each experimental article was rated in terms of several categories of independent variable assessment and independent variable definition. The occurrence or nonoccurrence of these discrete categories within each article was the variable of interest, and the data are reported in terms of the percentage of articles in which at least one independent variable falls into a category. Thus, for example, if a single article had three independent variables and reported accuracy assessment on all three variables, this article would be counted as a single occurrence of independent variable accuracy reporting. Similarly, if the article defined one of the variables but failed to define the others, the article would contribute once to the category of definition and once to the category of no definition.

Ratings of independent variable assessment were divided into three categories: (A) Yes, some form of assessment of the application of the independent variable was reported. This report could be informal (e.g., "observers agreed on all but one instance of the treatment variable application"), could indicate a statistical estimate of reliability between two observers (typically either percentage agreement or a correlational statistic) or could indicate calibration (a check by the experimenter to ensure that the actual occurrence matched the true value suggested by the method section). Selection of this category required clear evidence that at least one person had observed and recorded the occurrence of at least one treatment variable. (B) No, assessment was not reported but the application of the independent variable was judged to be at low risk for inaccuracy. This category included Kelly's (1977) definitions of mechanically defined treatments (e.g., a machine that delivered a token each time the subject pressed a button) and permanent products (e.g., the experimenter painted a garbage can with school colors). In addition, this category included single behavioral interventions (e.g., the experimenter gave one set of instructions or put up a sign) or continuous application of the independent variable (e.g., each time the subject turned in a piece of garbage, he received a token). It is conceivable that an experimenter might neglect to administer a reinforcer on a one for one basis, and that error could be present in the count of permanent products or in machine delivery, but lack of accurate administration of such variables would be less likely than in other, more complex schedules (Johnston & Pennypacker, 1980) and errors would be less likely to be biased (that is, to change bidirectionally with behavior). Thus, this category was used as a conservative approach to noting treatment variables in which error was possible but less likely than in category C. (C) No, independent variable accuracy checks were not reported and they were necessary. That is, the administration of the independent variable was not exempted by any of the cases cited in category B, and the potential for error was judged to be high.

The occurrence of independent variable operational definition was similarly divided into three categories: (A) Yes, an explicit operational definition was included. (B) No, an operational definition was not included but it was unnecessary. When the treatment variable was mechanical (e.g., the light turned on or a machine gave the child one M&M), very simplistic (e.g., the experimenter wrote on the board), had been well defined previously (e.g., the experimenter "modeled" a verbal response), or had a citation to a source describing it, it was included in this category. (C) No, an operational definition was not included and it was necessary, as it was not exempted by any of the cases cited in category B. This category contained treatment variables that were almost exclusively behaviors emitted by the treatment agent; typical variables included "praise," "affection," "positive social behaviors," and "chatting." Again, only independent variables clearly requiring definition were included in category C. If the major components of the independent variable were defined but a minor component (a multicomponent study using three types of tangible reinforcers, all accompanied by "praise") was not, the variable was categorized in A or B, not in C, in order to produce a conservative estimate of the need for treatment definition.

RESULTS

Coding Reliability

The reliability of observation of both independent variable assessment and definition of the independent variable was calculated by comparing the prime observer's ratings to a second observer who rated one issue per year from 1968 to 1980. Reasonable levels of agreement were obtained (Cohen's 1960, Kappa: $K = .80$, $K = .82$, respectively).

Independent Variable Reliability

Table 1 shows the number of experimental articles published each year. Figure 1 shows the percentage of articles reporting each of the three classes of independent variable assessment. These percentages can sum to greater than 100% since papers commonly used more than one independent variable or reported more than one experiment. In 1968, for example, 68% of the articles did not report assessment when the risk of inaccuracy was high for at least one independent variable, 32% of the articles did not report independent variable assessment when the risk of inaccuracy was low, and 23% reported assessment for at least one independent variable. As can be seen in Figure 1, the majority of articles did not report independent variable assessment even when the risk of in-

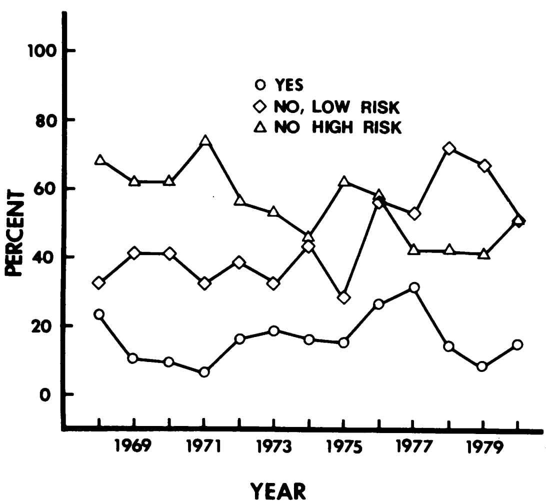

Fig. 1. Percentage of articles either presenting independent variable reliability where needed, not presenting independent variable reliability where risk of inaccuracy is high, or not presenting independent variable reliability where risk of inaccuracy is low. Percentages can sum to more than 100% because of multitreatment and multiexperiment articles.

Table 1

Number of Experimental Articles

Year	Volume	Articles
1968	1	31
1969	2	29
1970	3	34
1971	4	34
1972	5	45
1973	6	57
1974	7	56
1975	8	39
1976	9	43
1977	10	55
1978	11	36
1979	12	39
1980	13(1-3)	41

accuracy was high and there is little, if any, increase in reporting independent variable reliability over the past 12 years.

Independent Variable Operational Definition

Figure 2 shows the percentage of articles including operational definition of the independent variable. As can be seen, the majority of articles did report independent variable operational definitions when needed, although in a sizable number of cases (10%-50%) operational

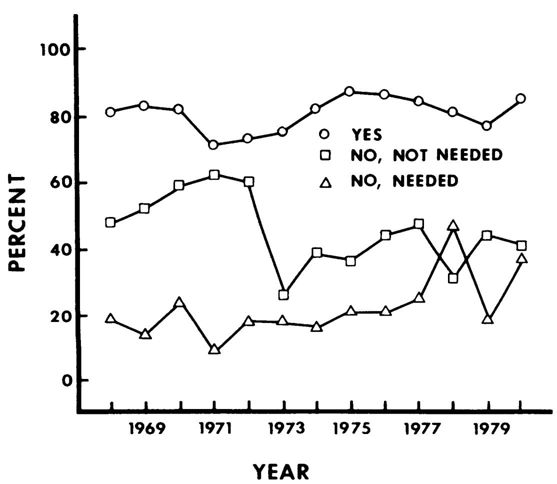

Fig. 2. Percentage of articles providing independent variable operational definitions where needed, not providing operational definitions where needed, or using independent variables not needing operational definitions. Percentages can sum to more than 100% because of multitreatment and multiexperiment articles.

definitions were not presented when necessary. There are no stable changes across time in these data, with the exception of a slight decrease in the number of independent variables not requiring operational definition in recent years.

The data on independent variable definition were collected both because definitions are often necessary for the evaluation and replication of experimental results and because the need for an operational definition can indicate the need for accuracy checks to see if the variable was used as defined. Among the surveyed studies presenting operational definitions, only an average of 16% (range 3-34%) also performed some check on the accuracy of the implementation of the independent variable.

DISCUSSION

The data document that the majority of articles published in *JABA* do not use necessary assessment of the independent variable. In addition, in a sizable minority of cases the independent variable is not operationally defined. Even when the independent variable is defined, in most cases no accuracy checks are made to see that it is used as defined. The data do not suggest any improvement in independent variable methodology since the journal was founded in 1968. Thus, the potential cost described previously would appear to be relevant to a large portion of the operant literature.

This review of the literature also revealed that there was more than one dimension of independent variable application in which inaccuracy could occur, although these were not rated separately. Inaccuracies in both the temporal and the physical dimensions of the independent variable application can be costly to the conclusions drawn from a study. For example, the temporal dimension refers to the appropriate occurrence of the independent variable in time in the ABAB or ABAC reversal sequence or in different portions of the multiple baseline design. The absence of overlap between different independent variables is important to demonstrate here. For example, if the design suggests that a parent or teacher who has been inadventently reinforcing a maladaptive response in baseline should ignore that response and reinforce alternative responses in treatment, it would be useful to identify treatment overlap reliably; that is, to show how often the maladaptive response was ignored in baseline and inadvertently reinforced in treatment. Particularly when reversals are unsuccessfully attempted, it is important to demonstrate the absence of the independent or treatment variable in the return to baseline phase. If no permanent record of the application of the independent variable exists, the experimenter may not be able to identify or exempt inaccurate treatment application from the list of causes of failure to find a functional relationship.

Another index of temporal reliability is the degree to which the treatment variable is delivered according to the schedule outlined in the method section. Since there is clear evidence with animals and some evidence with humans (e.g., Homer & Peterson, 1980) that changes in the scheduling of the independent variable can have a profound impact on responding, evidence concerning the actual delivery of reinforcement is necessary to avoid erroneous conclusions. For example, imagine a relatively untrained treatment agent applying a DRO 2-min schedule. The treatment agent begins by reliably checking his watch and delivering a food reinforcer contingent on "no tantrums" every 2 min on the dot. However, some tantrums continue and the therapist (who may have more experience with treating children than with using schedules) can tell when the child is about to have a tantrum. So he looks at his watch and says "1½ minutes, close enough" and delivers the reinforcer. Later, when the rate of tantrums is much lower, however, he may even forget a reinforcer or get it in late. When no data are collected on the treatment agent's behavior, it would not be apparent to the data collectors that any departure from the agreed on schedule had been made. Yet, the treatment agent has changed from a fixed DRO 2-min to an escalating DRO schedule and there are well-known differential effects for these two schedules (Homer & Peterson, 1980).

The physical dimension of independent variable reliability refers simply to what the treatment agent actually does, or the degree to which the description of the treatment in the method section matches the actual occurrence of treatment in practice. Again, many treatment agents are highly motivated to effect changes in behavior and may inadvertently begin to use additional treatment methods or may alter the current method to change or hasten results. For example, what if the treatment agent described above began to smile and wink at the child between reinforcers in order to maintain the withholding of tantrums? In this case, both facial expressions and food might serve as reinforcers and the actual schedule might be a DRO 15-sec rather than a DRO 2-min. Such treatment might be more likely to generalize to settings where smiles but not food was present. Similarly, the treatment agent might slowly shake his head and frown (signals for punishment in other settings) when the child had a tantrum, and thus, if effective, the actual schedule would be DRO + punishment. The above might also result in the conclusion that DRO was ineffective, if frowning served as an attentional reinforcer for tantrums. If naive observers are used, or even if experienced observers are sitting behind the treatment agent for a better view of the child, the physical reliability of the procedure cannot be determined, and the chance for results that could be reliably replicated is at risk. DeRubeis and Hollon (1981) in fact, suggest that both the therapist behaviors specified by the experimenter and the therapist-emitted behaviors not specified by the experimenter should be observed and reported as a measure of treatment variable accuracy.

Cost-Effective Solutions

Even if investigators agree that the potential for costly inaccuracy exists in independent variable methodology, there may be different reactions to this conclusion. The present authors, as well as some other investigators (e.g., Johnston & Pennypacker, 1980) feel that deliberately accepting inaccuracy in basic subject matter because it may be inconvenient to gather accurate information has no place in a science of behavior. Others may feel, however, that the potential cost of the inaccuracy must be balanced against the cost of ensuring accurate independent variable application. The cost-effectiveness of the solution would dictate whether it would be applied or not. If there were solutions which were cost-effective, then prevention of independent variable inaccuracy would be the most acceptable solution.

There are a variety of levels of solutions that might be suggested. At the lowest level of both cost and methodological completeness, this article might serve as a reminder to experimenters to ensure accurate independent variable application by rigorously training treatment agents and by periodic informal (and unreported) spot checks on the agent dispensing the independent variable. This solution might be a slight improvement on the status quo and would involve little additional cost. Since the potential for independent variable inaccuracy has rarely been explicitly discussed in the operant literature before, this discussion may spur any researchers not already engaged in such practices to more complete methodological safeguards. However, this solution leaves the possibility for continued inaccuracy. As Johnston and Pennypacker (1980) noted, when the complete description of all relevant information pertaining to a research study is not provided, the reader has only the choice of giving the author the benefit of the doubt or of making a more conservative assumption. Continuing to give the researcher the benefit of the doubt may leave a gap in acceptable behavior methodology.

A second solution might be to maintain the present practice of not assessing and reporting independent variable accuracy in cases where the risk of inaccuracy is low, but collecting and reporting data on independent variable application whenever the risk of inaccurate application is high. Thus, the reader would not be required to give the author the benefit of the doubt in cases where the doubt might be too liberal, and the researcher would retain the right to argue for or against the necessity of supplying such data. This would reduce the risk of inaccuracies in conclusions drawn from operant research at relatively low cost but it would not eliminate the risk.

A methodologically more conservative solution might be to use some measure of the accuracy of independent variable implementation routinely in much the same way the accuracy of dependent variable observation has been routinely examined, either by measuring independent variable reliability with multiple observers or by calibrating the observers against "true values" of the independent variable as specified by the researcher. If the researcher opts to take measures of independent variable reliability, the observers can simply be trained to observe the treatment variable as well as the dependent variable. Past data show that multiple variables can be reliably observed at the same time (e.g., Kent, O'Leary, Dietz, & Diament, 1979) and that data can be collected on two interacting individuals reliably (e.g., Pinkston, Reese, LeBlanc, & Baer, 1973). Some investigators have successfully observed and reported data both for the dependent and independent variables (e.g., Kazdin, 1977*b*). Simple occurrence/nonoccurrence of a crucial treatment variable can be observed in successive intervals to determine temporal reliability. Physical reliability could be determined simply by observing a category of "other therapist interaction," and then deciding whether to measure or control frequently occurring other behavior. The presentation of independent variable reliability would be a minor inconvenience for experimenters and could result in large improvements in the quality of experimental data.

Johnston and Pennypacker (1980) argue, however, that using multiple observers may not yield data relevant either to reliability or accuracy. Reliability typically refers to the extent to which an observation could be repeated to yield the same measure again, while accuracy refers to the extent to which the measure approximates the true value in nature. Simply because multiple observers agree on an observation does not automatically prove that the observation is reliable in a classic sense, since all the observers may share a bias or limitation which might systematically influence their measurement, reducing the extent to which the observation could be repeated successfully by a neutral source. Thus, simple observational reliability might not improve the actual reliability of independent variable implementation any more than multiple observer reliability can contribute conclusive information on the accuracy of dependent variable observation (Johnston & Pennypacker, 1980). A superior technique might be the method of calibration mentioned earlier, in which the data of the observer are compared with the true value in nature and inaccuracies in the observed values are then corrected. With the dependent variable, no "true" values are known, and Johnston and Pennypacker (1980) suggest a variety of methods of calibrating observers of the dependent variable including the use of bogus subjects who produce behaviors at a prearranged rate or the comparison of observed data to a permanent product or a mechanical record. Although the true value in nature of the independent variable is also an unknown, the methodology section of the research report suggests a known value that should (if the method section description is accurate) closely approximate the true value. Thus, unlike the dependent variable, on which no projection of the true value in nature is directly possible, the independent variable's true value in nature may be assumed to be the value specified by the experimenter. If the experimenter trains the observer to observe and record both independent and dependent variables, then the value of the observed independent variable can be compared to the specified value. Any difference between these values must be ascertained by the experimenter to be observational error or therapist drift. This calibration can be accomplished in the same way as dependent variable calibration, described in more detail by Johnston and Pennypacker (1980). This last method would appear to give firm assurance as to the accuracy of independent variable implementation.

Of course, there are many other cost-effective solutions to the problem of independent variable accuracy that may be used in prescribed situations. For example, correctly calibrated, mechanical methods of delivering or recording the occurrence of the independent variable which are equally effective as using human therapists and observers, might be preferred because they provide greater accuracy at lower cost. Similarly, program evaluators have suggested the routine application of process evaluation, which may consist of various means of documenting the occurrence of treatment variables (Donabedian, 1966; Way et al., 1978). Cook and Campbell (1979) suggested a variety of designs that may control for therapist drift in field settings. Finally, future investigators may engineer new, cost-effective methods to guarantee accurate independent variable application.

Thus, the observational technology used by behavior analysts suggests the need to assess the integrity of their reports of both the independent variable and the dependent variable. There are both published and anecdotal data showing that inaccuracies do occur. There is a clear potential cost for such inaccuracies, with the risk of failure to replicate and lack of generality of results being proportional to the degree of the inaccuracy. Finally, there are several cost-effective solutions to remove the risk, with perhaps the best solution being calibration of the treatment agent. The intent of the present article is not to lay down a new set of rigid requirements for behavior analysts; such rules would be antithetical to the practice of applied behavior analysis (Johnston & Pennypacker, 1980). Instead, the intent is to point out this potential methodological pitfall and to suggest some alternative strategies that may be used flexibly by behavior analysts to improve the quality of their data and the integrity of their conclusions regarding the effects of independent variables.

REFERENCE NOTES

1. Hollon, S. D., Mandell, M., Bemis, K. M., DeRubeis, R. J., Emerson, M., Evans, M. D., & Kriss, M. R. *Reliability and validity of the Young Cognitive Therapy Scale.* Unpublished manuscript, University of Minnesota, 1981.
2. DeRubeis, R. J. Personal communication, September 21, 1981.

REFERENCES

Arrington, R. E. Time sampling in studies of social behavior: A critical review of techniques and results with research suggestions. *Psychological Bulletin,* 1943, **40**, 81-124.

Baer, D. M., Wolf, M. M., & Risley, T. R. Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis,* 1968, **1**, 91-97.

Becker, J., & Schuckit, M. A. The comparative efficacy of cognitive therapy and pharmacotherapy in the treatment of depression. *Cognitive Therapy and Research,* 1978, **2**, 193-197.

Bellack, A. S., Hersen, M., & Himmelhock, J. M. Social skills training for depression: A treatment manual. *JSAS Catalog of Selected Documents,* 1980, **10**, 25. (MS. 2156).

Bernal, M. E., Klinnert, M. D., & Schultz, L. A. Outcome evaluation of behavioral parent training and client-centered parent counseling for children with conduct problems. *Journal of Applied Behavior Analysis,* 1980, **13**, 677-691.

Bijou, S. W., Peterson, R. F., & Ault, M. H. A method to integrate descriptive and experimental field studies at the level of data and empirical concepts. *Journal of Applied Behavior Analysis,* 1968, **1**, 175-191.

Billingsley, F., White, O. R., & Munson, R. Procedural reliability: A rationale and an example. *Behavioral Assessment,* 1980, **2**, 229-241.

Bornstein, P., & Quevillon, R. Effects of a self-instructional package on overactive preschool boys. *Journal of Applied Behavior Analysis,* 1976, **9**, 179-188.

Carnine, D. W. Effects of two teacher-presentation rates on off-task behavior, answering correctly, and participation. *Journal of Applied Behavior Analysis,* 1976, **9**, 199-206.

Cohen, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement,* 1960, **20**, 37-46.

Cook, T. D., & Campbell, D. T. *Quasi-experimentation: Design and analysis issues for field settings.* Chicago: Rand McNally, 1979.

Dallemagne, G., & Richelle, M. Titration schedule with rats in a restraining device. *Journal of Experimental Analysis of Behavior,* 1970, **13**, 339-348.

DeRubeis, R. J., & Hollon, S. D. Behavior treatment of affective disorders. In L. Michelson, M. Hersen, & S. M. Turner (Eds.), *Future perspectives in behavior therapy.* New York: Plenum, 1981.

Donabedian, A. Evaluating the quality of medical care. *Milbank Memorial Fund Quarterly,* 1966, **44**, 166-203.

Dorsey, M. F., Iwata, B. A., Ong, P., & McSween, T. E. Treatment of self-injurious behavior using a water mist: Initial response suppression and generalization. *Journal of Applied Behavior Analysis,* 1980, **13**, 343-353.

Fleischman, M. J. A replication of Patterson's "Intervention for boys with conduct problems." *Journal of Consulting and Clinical Psychology,* 1981, **49**, 342-351.

Friedling, C., & O'Leary, S. G. Effects of self-instruction training on second- and third-grade hyperactive children: A failure to replicate. *Journal of Applied Behavior Analysis,* 1979, **12**, 211-219.

Hanley, E. M. Review of research involving applied behavior analysis in the classroom. *Review of Educational Research,* 1970, **40**, 597-625.

Hartmann, D. P. Considerations in the choice of interobserver reliability estimates. *Journal of Applied Behavior Analysis,* 1977, **10**, 103-116.

Hartmann, D. P., Roper, B. L., & Gelfand, D. M. An evaluation of alternative modes of child psychotherapy. In B. B. Lahey & A. E. Kazdin (Eds.), *Advances in clinical child psychology,* (Vol. 1). New York: Plenum Press, 1977.

Hasazi, J. E., & Hasazi, S. E. Effects of teacher attention on digit-reversal behavior in an elementary school child. *Journal of Applied Behavior Analysis,* 1972, **5**, 157-162.

Hayes, S. C., Rincover, A., & Solnick, J. V. The technical drift of applied behavior analysis. *Journal of Applied Behavior Analysis,* 1980, **13**, 275-285.

Hersen, M. Complex problems require complex solutions. *Behavior Therapy,* 1981, **12**, 15-29.

Hersen, M., & Barlow, D. H. *Single case experimental designs.* New York: Pergamon, 1976.

Heyns, R. W., & Lippitt, R. Systematic observational techniques. In G. Lindzey (Ed.), *Handbook of social psychology,* Vol. 1. Cambridge, Mass.: Addison-Wesley, 1954.

Hobbs, S. A., Moguin, L. E., Tyroler, M., & Lahey, B. B. Cognitive behavior therapy with children: Has clinical utility been demonstrated? *Psychological Bulletin,* 1980, **87,** 147-165.

Homer, A. L., & Peterson, L. Differential reinforcement of other behavior: A preferred response elimination procedure. *Behavior Therapy,* 1980, **11,** 449-471.

Johnston, J., & Pennypacker, H. S. *Strategies and tactics of human behavioral research.* Hillsdale, N.J.: Erlbaum, 1980.

Kazdin, A. E. The effect of vicarious reinforcement on attentive behavior in the classroom. *Journal of Applied Behavior Analysis,* 1973, **6,** 71-78.

Kazdin, A. E. Artifact, bias, and complexity of assessment: The ABCs of reliability. *Journal of Applied Behavior Analysis,* 1977, **10,** 141-150. (*a*)

Kazdin, A. E. The influence of behavior preceding a reinforced response on behavior change in the classroom. *Journal of Applied Behavior Analysis,* 1977, **10,** 299-310. (*b*)

Kelly, M. B. A review of the observational data-collection and reliability procedures reported in *The Journal of Applied Behavior Analysis. Journal of Applied Behavior Analysis,* 1977, **10,** 97-101.

Kent, R. N., O'Leary, K. D., Dietz, A., & Diament, C. Comparison of observational recordings in vivo, via mirror, and via television. *Journal of Applied Behavior Analysis,* 1979, **12,** 517-522.

Mathews, R. M., & Fawcett, S. B. Community applications of instructional technology: Training low-income proctors. *Journal of Applied Behavior Analysis,* 1977, **10,** 747-754.

McNamara, J. R., & MacDonough, T. S. Some methodological considerations in the design and implementation of behavior therapy research. *Behavior Therapy,* 1972, **3,** 361-378.

Milby, J. B. Modification of extreme social isolation by contingent social reinforcement. *Journal of Applied Behavior Analysis,* 1970, **3,** 149-152.

Nelson, G. L., & Cone, J. D. Multiple-baseline analysis of a token economy for psychiatric inpatients. *Journal of Applied Behavior Analysis,* 1979, **12,** 255-271.

O'Leary, K. D., & Kent, R. N. Behavior modification for social action: Research tactics and problems. In L. A. Hamerlynck, L. C. Handy, & E. J. Mash (Eds.), *Behavior change: Methodology, concepts, and practice.* Champaign, Ill.: Research Press, 1973.

O'Leary, K. D., Kent, R. N., & Kanowitz, J. Shaping data collection congruent with experimental hypotheses. *Journal of Applied Behavior Analysis,* 1975, **8,** 43-52.

Patterson, G. R. Interventions for boys with conduct problems: Multiple settings, treatments, and criteria. *Journal of Consulting and Clinical Psychology,* 1974, **42,** 471-481.

Paul, G. L., & Lentz, R. J. *Psychological treatment for chronic mental patients: Milieu versus social learning programs.* Cambridge, Mass.: Harvard University Press, 1977.

Pinkston, E. M., Reese, N. M., LeBlanc, J. M., & Baer, D. M. Independent control of a preschool child's aggression and peer interaction by contingent teacher attention. *Journal of Applied Behavior Analysis,* 1973, **6,** 115-124.

Porterfield, J. K., Herbert-Jackson, E., & Risley, T. R. Contingent observation: An effective and acceptable procedure for reducing disruptive behavior of young children in a group setting. *Journal of Applied Behavior Analysis,* 1976, **9,** 55-64.

Reid, J. B. Reliability assessment of observational data: A possible methodological problem. *Child Development,* 1970, **41,** 1143-1150.

Reisinger, J. J. The treatment of "anxiety-depression" via positive reinforcement and response cost. *Journal of Applied Behavior Analysis,* 1972, **5,** 125-130.

Romanczyk, R. G., Kent, R. N., Diament, C., & O'Leary, K. D. Measuring the reliability of observational data: A reactive process. *Journal of Applied Behavior Analysis,* 1973, **6,** 175-184.

Rosenthal, R., & Rosnow, R. L. *Artifact in behavioral research.* New York: Academic Press, 1969.

Rounsaville, B. J., Weissman, M. M., & Prusoff, B. A. Psychotherapy with depressed outpatients: Patient and process variables as predictors of outcome. *British Journal of Psychiatry,* 1981, **138,** 67-74.

Sidman, M. *Tactics of scientific research.* New York: Basic Books, 1960.

Stallings, J. Implementation and child effects of teaching practices in Follow Through classrooms. *Monographs of the Society for Research in Child Development,* 1975, **40,** Nos. 7-8.

Strain, P. S., Shores, R. E., & Timm, M. A. Effects of peer social initiations on the behavior of withdrawn preschool children. *Journal of Applied Behavior Analysis,* 1977, **10,** 289-298.

Urbain, E. S., & Kendall, P. C. Review of social-cognitive problem-solving interventions with children. *Psychological Bulletin,* 1980, **88,** 109-143.

Way, J. R., Lund, D. A., & Attkisson, C. C. Quality assurance in human service program evaluation. In C. C. Attkisson, W. A. Hargreaves, M. J. Horowitz, & J. E. Sorensen (Eds.), *Evaluation of human service programs.* New York: Academic Press, 1978.

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrist, L. *Unobtrusive measures: Nonreactive* *research in the social sciences.* Chicago: Rand McNally, 1966.

Weick, K. E. Systematic observational methods. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology,* Vol. 2. Menlo Park, Calif: Addison-Wesley, 1968.

Wodarski, J. S., Feldman, R. A., & Pedi, S. J. Objective measurement of the independent variable: A neglected methodological aspect in community-based behavioral research. *Journal of Abnormal Child Psychology,* 1974, **2**, 239-244.

Received October 26, 1981
Final acceptance April 6, 1982