

Several of the current methodological problems for behavior modification research and practice are considered to be a direct outcome of recent developments of the field as a therapeutic-educational endeavor. In order to highlight these problems, behavioral intervention strategies are considered across the sequential phases in the life cycle of a treatment method presented by Bachrach and Quigley (1966): case-study, comparative study, and follow-up study. Within each of these phases the types of treatment questions being asked are shown to be a prime factor in the kinds of methodological concerns which have arisen. These methodological concerns include those related to the selection of measuring instruments, design strategies, and implications for data analysis.

ERIC J. MASH*

Behavior Modification and Methodology: A Developmental Perspective

Proponents of behavior modification are supposedly "methodology conscious." Indeed, an emphasis on sound data has often been specified as a distinguishing feature of learning-based treatment approaches in contrast to approaches based on more dynamic constructs.¹ The common-law marriage of behavior modification to science, the legitimacy of which has been questioned by some,² has produced the miscegenations of researcher/practitioner and/or practitioner-researcher.³ Given their common background, these offspring should presumably exhibit a marked concern for the soundness of data. But in spite of the common lineage and predisposing influences, the progenies have been somewhat of a methodological disappointment. Several major reviews of behavior modification studies have concluded on an ominous note.⁴ With respect to a number of criteria for evaluating research methodology —

*Eric J. Mash is Associate Professor of psychology, The University of Calgary, Calgary, Alberta, Canada. During the writing of this paper the author was supported by Canada Council grants S72-0461 and S70-1567.

¹S. M. Johnson and O. D. Bolstad, ["Methodological Issues in Naturalistic Observation: Some Problems and Solutions for Field Research," in L. A. Hamerlynck, L. C. Handy, and E. J. Mash (eds.), *Behavior Change: Methodology, Concepts and Practice* (Champaign, Ill.; Research Press, 1973); E. A. Ramp and B. L. Hopkins, *A New Direction for Education: Behavior Analysis, 1971* (Kansas: Department of Human Development, 1971); *Case Studies in Behavior Modification*, ed. L. P. Ullman and L. Krasner (New York: Holt, Rinehardt & Winston, 1965).

²G. C. Davison and S. J. Taffel, "Effects of Behavior Therapy," a paper presented at the American Psychological Association Convention, Honolulu, Hawaii, 1972; P. London, "The End of Ideology in Behavior Modification," *American Psychologist*, 27 (1972), 913-920; E. J. Mash, "Has Behavior Modification Lost Its Identity?," *Canadian Psychologist*, 15 (1974), 271-280.

³S. Axelrod, "Education and Science: Compatible Endeavors," in *A New Direction for Education: Behavior Analysis 1971*, ed. E. A. Ramp and B. L. Hopkins (Kansas: Department of Human Development, 1971); T. R. Risley, "Behavior Modification: An Experimental-Therapeutic Endeavor," in *Behavior Modification and Ideal Mental Health Services*, ed. L. A. Hamerlynck, P. O. Davidson, and L. E. Acker (Calgary: John D. McAra, 1969).

⁴L. Breger and J. L. McGaugh, "Critique and Reformulation of 'Learning Theory' Approaches to Psychotherapy and Neurosis," *Psychological Bulletin*, 63 (1965), 338-358; D. M. Gelfand and D. P. Hartmann, "Behavior Therapy with Children: A Review and Evaluation of Research Methodology," *Psychological Bulletin*, 69 (1968), 204-215; R. Pawlicki, "Behavior-Therapy Research with Children: A Critical Review," *Canadian Journal of Behavioral Science*, 2 (1970), 165-173.

criteria which includes both issues of data collection and experimental design — a large proportion of the studies reviewed were seen as being deficient. Several more recent papers⁵ suggest that since the time of these reviews the situation has not changed much, although in these papers there is also an expression of a belief that many of the methodological deficiencies can be corrected, as well as constructive offerings towards this end.

It is not so unusual for “thinkers” in an area to eschew questions of methodology. Matters of method may be relegated to secondary importance superseded by more basic questions of theoretical import. Oftentimes method is assumed, or set aside for the technician. Such thinking is clearly erroneous. It is erroneous in that it is based upon the supposition that methodological queries address themselves only to technique; that they are questions about the adequacy of a design paradigm, a measuring instrument (human or otherwise), a way of summarizing or analyzing data, etc. It is critical to recognize that questions subsumed under methodology extend well beyond technique. For example, regarding generalizability, Cronbach and his associates have made the following point: “Questions concerning generalizability are substantive, not purely methodological. They inquire about the postulated class of observations as well as about the measuring technique.”⁶

In many clinically-oriented endeavors it is not uncommon for the development of change procedures to take initial precedence over systematic observation and careful methodology.⁷ It is the contention of this paper that many of the current methodological problems and developments in behavior modification in many ways stem from the development of the field as a therapeutic-educational endeavor. A brief description of this course of development may help to place in an historical perspective, some of the methodological issues which have recently and increasingly been given a greater amount of attention.

Bachrach and Quigley have characterized almost all therapeutic methods as having a fairly typical “life cycle.” They state:

There is, however, a similar life cycle in all such applications, which runs as follows: First, after the initial proposal of the usefulness of a particular method — let's stick to therapeutic techniques for this — there is a high frequency of published papers and verbal reports (at professional meetings, for example) on clinical applications of the therapy. These papers and reports are almost inevitably of the case-study type — an empirical report on method X used on *n* number of patients with certain results. Following the case-study phase of the life cycle are a drop in the frequency of the case reports and an increase in the studies in which method X is compared with other methods . . . in terms of clinical effectiveness. This *comparison* phase leads into the next, or *follow-up* stage, in which other reports (presented with much

⁵D. H. Barlow and M. Hersen, “Single-Case Experimental Designs,” *Archives of General Psychiatry*, 29 (1973), 319-325; Johnson and Bolstad, “Methodological Issues,” pp. 7-67; A. E. Kazdin, “Methodological and Assessment Considerations in Evaluating Reinforcement Programs in Applied Settings,” *Journal of Applied Behavior Analysis*, 6 (1973), 517-531; D. Lipinski and R. Nelson, “Problems in the Use of Naturalistic Observations as a Means of Behavioral Assessment,” *Behavior Therapy*, 5 (1974), 341-351; K. D. O’Leary and R. Kent, “Behavior Modification for Social Action: Research Factors and Problems,” in *Behavior Change: Methodology, Concepts, and Practice*, ed. L. A. Hamerlynck, L. C. Handy, and E. J. Mash (Champaign, Ill.: Research Press, 1973).

⁶L. J. Cronbach, N. Rajaratnam, and G. C. Gleser, “Theory of Generalizability: A Liberalization of Reliability Theory,” *British Journal of Statistical Psychology*, 16 (1963), 137-163.

⁷R. Q. Bell, “Structuring Parent-Child Interaction Situations for Direct Observation,” *Child Development*, 35 (1964), 1009-1020; A. M. Bodin, “Conjoint Family Assessment: An Evolving Field,” in *Advances in Psychology Assessments*, ed. P. McReynolds (Palo Alto: Science and Behavior Books, 1968).

less frequency than the other types) appear, bringing the long range results of therapy to the light: . . . The final phase of the life cycle of the technique is the re-evaluative phase, in which the mode of action of the method is explored . . . In sum, the sequential phases of *case study*, *comparison*, *follow-up*, and *mode of action* may portray the life cycle of the method.⁸

With particular reference to behavior modification they go on to say:

Applying this to the behavior therapies, we find that these techniques are, for the most part, in the first phase, that of case reports. For example, Ullmann and Krasner's book called *Case Studies in Behavior Modification* (1965) consists solely of case studies. Their other work (Krasner & Ullmann, 1965) called *Research in Behavior Modification* offers some intensive papers on research, in which much of the material is entirely based on case reports. This is perfectly sound practice; what is now probable is an increase in the frequency of such case studies for a period, followed (if our life cycle holds true) by a drop in case reports and an increase in comparative studies (not necessarily done by the practitioners themselves — behavior modification versus psychoanalytic therapy or nondirective therapy). The follow-ups and, ultimately, the mode of action in which such concepts as response and reinforcement will have to be considered carefully from a theoretical standpoint will be discussed in greater detail later. Application in the field always is ahead of the research that tests it — clinical exigencies may demand this — but application divorced from research and theory has limited utility.⁹

Considering these remarks eight years later, the statements by Bachrach and Quigley seem prophetic. Behavior modification is now barely emerging from its embryonic (case-study) stage. For example, Patterson states:

. . . we are now ready for phase two, which will be characterized by several concerns. While continuing to use operational language, observation data, and a functional analysis, phase two will also become more analogous to social engineering. This implies a consideration of the *efficiency* of intervention approaches. How effective is a given technique when applied to *consecutive* cases, all of which are characterized by similar problems . . . The third characteristic will be systematic collection of follow-up data. Procedures which produce changes in behavior, at reasonable cost but with low maintenance, have a short-term efficiency but may not serve society's needs because of the temporary nature of the effect.¹⁰

He goes on to say:

Certainly, behavior modification *per se*, is not some monolithic system which generates, deductively, a single intervention approach tailored to each problem. The fact that several different approaches could be designed to alter the same class of behaviors suggests the necessity for comparative studies.¹¹

Clearly, Patterson's statements reflect the earlier remarks by Bachrach and Quigley, and would suggest that the 1970s marked the beginning of stage two for behavior modification.

Given that such a life cycle process is operative, the question can be raised as to how the development of behavior modification as a treatment approach relates to, and interacts with, methodological concerns. This question will now be discussed in relation to some of the recent measurement and design problems that have received attention from behavioral researchers. Methodological developments will be considered across the sequential stages; case studies, comparative studies and follow-up studies.

⁸A. J. Bachrach and W. A. Quigley, "Direct Methods of Treatment," in *Introduction to Clinical Psychology*, ed. I. A. Berg and L. A. Pennington (New York: Ronald Press, 1966).

⁹*Ibid.*, p. 509.

¹⁰G. R. Patterson, "Intervention in the Homes of Predelinquent Boys: Steps Toward Stage Two," a paper prepared for the workshop, *Delinquent Behavior: Some Psychological Research and Applications*, at the American Psychological Association Convention, Washington, D.C., 1971.

¹¹*Ibid.*, p. 2.

CASE STUDIES

With respect to the types of measuring instruments employed, developments during the case-study stage, may be directly attributed to an attempt at bridging the gap between measurement in the animal laboratory and measurement in social situations involving humans,¹² as well as a concomitant shifting emphasis from operant learning to social learning theory.

Animal Laboratory to Human Social Situations

Early case-studies were typically of the single-behaviour, single-subject variety.¹³ The measuring instruments which were used in these studies were quite simple, as the dependent measures were typically a frequency count of a particular behavior which was often later converted to a rate measure. Thus, the measuring instrument was essentially a behavior checklist¹⁴ perhaps supplemented by the use of a counter¹⁵ and/or some type of timing device.¹⁶ For all practical purposes, the measuring instrument was not much different from the electro-mechanical monitoring devices of the operant conditioning chamber used to study the behavior of rats and pigeons — with at least three major exceptions: The new subjects for study (people) exhibited a *multiple response* repertoire, of *greater complexity*, and they also exhibited more *mobility*. These exceptions necessitated several major changes in the measuring instruments.

The first exception involved an increase in the number of behaviors observed. While it was feasible to consider the response repertoire of a pigeon as consisting of a single response and no others (although many stalwart cleaners of operant conditioning chambers probably would have argued with this contention!), behavior modifiers found themselves concerned with more than one response.¹⁷ The recognition that there would likely be several behaviors of major interest meant the use of measuring instruments that contained multiple responses and also created the problem of *which* behavioral attribute(s) should be observed and measured.¹⁸

The greater qualitative complexity of the behaviors being observed also created methodological problems. Often, the response definition could no

¹²D. M. Baer, M. M. Wolf, and T. R. Risley, "Some Current Dimensions of Applied Behavior Analysis," *Journal of Applied Behavior Analysis*, 1 (1968), 91-97; C. B. Ferster, "Transition from Animal Laboratory to Clinic," *Psychological Record*, 17 (1967), 145-150.

¹³Ullman and Krasner, *Case Studies*.

¹⁴B. H. Hart *et al.*, "Effects of Social Reinforcement on Operant Crying," *Journal of Experimental Child Psychology*, 1 (1964), 145-153.

¹⁵O. R. Lindsley, "A Reliable Wrist Counter for Recording Behavior Rates," *Journal of Applied Behavior Analysis*, 1 (1968), 77-78; R. L. Mattos, "A Manual Counter for Recording Multiple Behavior," *Journal of Applied Behavior Analysis*, 1 (1968), 130.

¹⁶R. C. Worthy, "A Miniature, Portable Timer and Audible Signal-Generating Device," *Journal of Applied Behavior Analysis*, 1 (1968), 159-160.

¹⁷G. R. Patterson and G. D. Brodsky, "A Behavior Modification Programme for a Child with Multiple Behavior Problems," *Journal of Child Psychology and Psychiatry*, 7 (1966), 277-295; M. M. Wolf, T. Risley, and H. Mees, "Application of Operant Conditioning Procedures to the Behavior Problems of an Autistic Child," *Behavior Research and Therapy*, 1 (1964), 305-312.

¹⁸L. R. Goldberg, "Some Recent Trends in Personality Assessment," invited lecture presented at the meetings of the American Psychological Association, Washington, D.C., 1971; H. Lytton, "Observation Studies of Parent-Child Interaction: A Methodological Review," *Child Development*, 42 (1971), 651-684; J. S. Wiggins, *Personality and Prediction: Principles of Personality Assessment* (Reading, Mass.: Addison-Wesley, 1972).

longer be made in terms of the closure of a micro-switch, and it became more difficult to define what the response unit was. For a time there was a tendency to rely upon molecular-operational definitions. So, for example, we have statements such as, "Thumbsucking was recorded on a Gerbrands cumulative recorder which stepped one response for every three cumulative seconds of thumbsucking."¹⁹ However, this created the difficulty that in an attempt to be operational (and get people to agree) it was necessary to break behaviors down into very small units. The fragmentation of human behavior into smaller bits resulted in a certain loss of meaning, and the types of measuring instruments which later appeared seemed to represent more of a molar compromise in response definition. So, we later find definitions such as, "PP (Physical positive): Use this category whenever a person touches another in a friendly or affectionate manner . . .,"²⁰ or "Interaction: An attempt to initiate or maintain some type of mutual contact . . ."²¹ The problem of definition also meant a shift away from a mechanical measuring device to an instrument with presumably greater discriminatory capability — the human observer — but also, with greater susceptibility to inaccuracy.²²

Contributing, also, to the need for human observers was the fact that mechanical measurement, with the possible exceptions of remote instrumentation²³ and telemetry,²⁴ was only possible with a captive audience; that is, the behavior had to be where the machine was and part of the response definition included some interaction with the machine. With mobile human subjects this was a more or less unnatural situation, and, in addition, interest was becoming more focussed on *social* behaviors, which made it difficult to include instrumentation as part of the response. Concerning this need for human observers, Baer, Wolf and Risley state, "Current applied research . . . also suggests that instrumental recording with its typical reliability is not always possible. The reliable use of human beings to quantify the behavior of other human beings is an area of psychological technology long since well developed, thoroughly relevant, and very often necessary to applied behavior analysis."²⁵

It is thus evident that early in the case study stage, seemingly minor changes in procedure necessitated by the shift to the application of behavioral analysis with higher organisms, produced some rather significant changes in the nature of the measuring instrument used. There were a much larger number of behaviors observed, of greater complexity, and reliance upon the human observer as a standard measurement device increased. These changes in measuring instrument interact with both design and data analysis. For

¹⁹D. M. Baer, "Laboratory Control of Thumbsucking by Withdrawal and Re-presentation of Reinforcement," *Journal of the Experimental Analysis of Behavior*, 5 (1962), 525-528.

²⁰G. R. Patterson *et al.*, A Manual for Coding of Family Interactions, 1969 Revision," unpublished manuscript, University of Oregon, 1969.

²¹E. J. Mash, L. Terdal, and K. Anderson, "The Response Class Matrix: A Procedure for Recording Parent-Child Interactions," *Journal of Consulting and Clinical Psychology*, 40 (1973), 163-164.

²²R. E. Arrington, "Some Technical Aspects of Observer Reliability as Indicated in Studies of the 'Talkies,'" *American Journal of Sociology*, 38 (1932), 409-417.

²³R. L. Schwitzgebel and R. M. Bird, "Sociotechnical Design Factors in Remote Instrumentation with Humans in Natural Environments," *Behavior Research Methods and Instrumentation*, 2 (1970), 99-105.

²⁴C. A. Caceres, ed., *Biomedical Telemetry* (New York: Academic Press, 1965).

²⁵Baer, Wolf and Risley, p. 93.

example, more categories of greater complexity in a system have major implications for the number of subjects to be observed,²⁶ the number of observers to be employed and the amount of time spent on observer training.²⁷ Given that more than one category is involved in a behavior rating system, questions then arise as to whether the categories are independent or correlated, as well as, whether or not the measuring device possesses ipsative or differential properties.²⁸ These questions, in turn, have important implications for data analyses.²⁹

Operant Psychology to Social Learning

Certain theoretical developments with respect to treatment during the case-study stage also produced rather significant changes in the type of measuring instruments employed. These developments involved primarily a shift from "operant-behavior modification" to "social-learning behavior modification." Social learning behavior modification as presented by both Bandura and Gewirtz represents an extension of the principles of operant psychology.³⁰ However, as has been pointed out, "while the principles are well established and understood, their application to practical problems of human engineering have, as yet, simply not been spelled out."³¹ The movement toward social learning treatment approaches provided a new treatment emphasis. Given the basic learning assumption that most human behaviors are learned and maintained by members of the immediate social environment, several factors become evident.

(1) The arena for observation of behavior becomes the natural environment — it is where the action is. Presumably, laboratory analogue situations which reflect interactions as they occur in the natural environment would also be appropriate.³² and sometimes necessary to obtain data relevant to the hypotheses being tested.³³ However, some of the difficulties involved in setting up

²⁶M. B. Thorne, R. S. Schlottman, and B. Seary, "Single Animal Observation Versus Pair Observation in Macca Iris," *Journal of Genetic Psychology*, 115 (1969), 17-32.

²⁷E. J. Mash and J. D. McElwee, "Situational Effects on Observer Accuracy: Behavioral Predictability, Prior Experience, and Complexity of Coding Categories," *Child Development*, 45 (1974), 367-377.

²⁸R. B. Cattell, "Psychological Measurement: Normative, Ipsative, Interactive," *Psychological Review*, 51 (1944), 292-303.

²⁹R. R. Jones, "Behavioral Observation Frequency Data: Problems in Scoring, Analysis, and Interpretation," in *Behavior Change: Methodology, Concepts and Practice*, ed. L. A. Hamerlynck, L. C. Handy, and E. J. Mash (Champaign, Ill.: Research Press, 1973), pp. 119-145.

³⁰A. Bandura, *Principles of Behavior Modification* (New York: Holt, Rinehart & Winston, 1969); J. L. Gewirtz, "Mechanisms of Social Learning: Some Roles of Stimulation and Behavior in Early Human Development," in D. A. Goslin (ed.), *Handbook of Socialization Theory and Research* (Chicago: Rand-McNally, 1963).

³¹G. R. Patterson, J. A. Cobb, and R. S. Ray, "A Social Engineering Technology for Retraining the Families of Aggressive Boys," *Issues and Trends in Behavior Therapy*, ed. H. Adams and I. P. Unikel (Springfield, Ill.: Charles C. Thomas, 1973) pp. 139-224.

³²D. A. Bernstein and G. L. Paul, "Some Comments on Therapy Analogue Research with Small Animal 'Phobias'," *Journal of Behavior Therapy and Experimental Psychiatry*, 2 (1971), 225-233; T. D. Borkovec et al.; "Evaluation of a Clinically Relevant Target Behavior for Analogue Outcome Research," *Behavior Therapy*, 5 (1974), 503-513; R. M. McFall and A. R. Marston, "An Experimental Investigation of Behavior Rehearsal in Assertive Training," *Journal of Abnormal Psychology*, 76 (1970), 295-303.

³³Bell, "Structuring Parent-Child Interaction," pp. 1009-1020.

such situations³⁴ have led people to view the natural environment — home, classroom, etc. — as the more direct route.³⁵

(2) The target for intervention becomes a social unit, rather than a single individual.

(3) The dependent measure may consist of at least a dyadic interaction rather than a single response in isolation.

(4) The temporal occurrence of behavior including both antecedent and consequent stimuli becomes important for a functional analysis.³⁶

(5) Since the reinforcers in the situation are of a social nature, the stimulus-response-reinforcement relationships become blurred, as the response of people are also stimuli for other responses, as well as, reinforcers for responses when presented in a consequent temporal relationship.³⁷

These deductions from a social-learning framework, have had a profound influence on the measuring instruments employed, with their concomitant methodological problems. First, the difficulties involved in the use of observational methodology in naturalistic settings have been discussed by a number of authors.³⁸ These discussions include some of the most recent empirical studies dealing with the use of human observers for the purpose of obtaining data in natural settings. The word "empirical" is to be emphasized here, as observations in natural settings have for a long time had a certain amount of folklore about them.³⁹ This folklore usually is concerned with the observer's influence on the situation and the statement that "they get used to us" has been reiterated over a long period of time with little systematic empirical support. In spite of the recent empirical coverage given to methodological problems of naturalistic observation within a behavioral framework⁴⁰ it should be

³⁴S. Martin, "The Comparability of Behavioral Data in Laboratory and Natural Settings," paper presented at the Banff International Conference on Behavior Modification, Banff, Alberta, Canada, 1974.

³⁵R. B. Tharp and R. J. Wetzel, *Behavior Modification in the Natural Environment* (New York: Academic Press, 1969); J. S. Wiggins, "The Quality of Observational Data: Discussion," paper presented at a symposium on the Quality of Observational Data, held at the annual meetings of the Western Psychological Association, San Francisco, April, 1974.

³⁶S. W. Bijou and R. F. Peterson, "The Psychological Assessment of Children: A Functional Analysis," in *Advances in Psychological Assessment*, Vol. 2, ed. P. McReynolds (Palo Alto, California: Science and Behavior Books, 1971), pp. 63-78.

³⁷G. R. Patterson and J. A. Cobb, "A Dyadic Analysis of 'Aggressive' Behaviors," in *Minnesota Symposia on Child Psychology*, Vol. 5, ed. J. P. Hill (Minneapolis: University of Minnesota, 1971), pp. 72-129.

³⁸S. J. Hutt and C. Hutt, *Direct Observation and Measurement of Behavior* (Illinois: Charles C. Thomas, 1970); Johnson and Bolstad, pp. 7-67; Lipinski and Nelson, pp. 341-351; *Behavior-Therapy Assessment: Diagnosis, Design and Evaluation*, E. J. Mash and L. G. Terdal (New York: Springer Publishing Company, Inc., 1976); E. P. Willems and H. L. Raush, *Naturalistic Viewpoints in Psychological Research* (New York: Holt, Rinehart & Winston, 1969); H. F. Wright, "Observational Child Study," in *Handbook of Research Methods in Child Development*, ed. P. H. Mussen (New York: Wiley, 1960); pp. 71-139.

³⁹D. S. Thomas, "An Attempt to Develop Precise Measurement in the Social Behavior Field," *Sociologist*, 9 (1933), 1-21.

⁴⁰Johnson and Bolstad, pp. 7-67; Mash and McElwee, "Situational Effects," pp. 367-377; O'Leary and Kent, pp. 69-96; J. B. Reid, "Reliability Assessment of Observation Data: A Possible Methodological Problem," *Child Development*, 41 (1970), 1143-1150; R. G. Romanczyk et al., "Measuring the Reliability of Observational Data: A Reactive Process," *Journal of Applied Behavior Analysis*, 6 (1973), 175-184.

recognized that the statements being made and the conclusions being drawn are based upon a somewhat restricted data base — primarily, families at home and children in classrooms and playrooms. The quality of data collected across innumerable other naturalistic settings (work, recreation, etc.) and across other populations (adults' problems) has not yet been investigated.

In addition, it is believed that to the extent that there is control of behavior and stimuli in a situation, by the investigator these situations must be viewed as "quasi-natural."⁴¹ These points are not necessarily made as criticisms, but rather as a suggestion for possible future directions for research. Thus, the movement of behavior modification from an "operant set" to a "social learning set" was accompanied by a movement towards naturalistic observation which, as described in a number of studies, produced numerous problems regarding reactivity of both the measuring instrument and the subject of measurement.

Second, the fact that the subject of interest was now a social unit compounded the complexity which initially involved an increase in the number of target behaviors for a single individual. Now, the dependent measure was not only several behaviors, but also several people. This placed increasing demands on the observer, both for training and observation. It also resulted in a shift in the types of response definitions employed. When a single individual was the target of observation it was possible to still be fairly idiosyncratic regarding definition. "Hitting" for Johnny might consist of a left jab or right uppercut. However, when Johnny was being observed as part of a social unit with five or more other children and/or adults, who threw right jabs and left uppercuts, the specific response topography described for Johnny was no longer applicable. It has not been feasible to maintain idiosyncratic response definitions for all persons being observed, since it seems to exhaust the information processing capabilities of the observer. The trend has been toward the development of response definitions in terms of categories or classes of responses.⁴² Hitting might now be defined as aggression, which would in turn represent quite distinct responses (kick versus slap) for two different individuals. In effect, topographically different responses might be labelled the same. Clearly, some degree of precision was lost with categorical measurement. For example, from a previous measurement we might predict that child X would be "physically aggressive" towards us at the rate of five times per minute; however, we would not know whether he was going to hit, kick or throw something at us. At the same time, there are benefits resulting from categorical description — a primary one being that it is possible to make *inter-individual comparisons*, a feature which is quite important to the comparative stage of the life cycle for a treatment method.

A critical point in the application of measuring instruments with categorical components relates to the manner in which the categories are formed. Thus far category definition has been a somewhat arbitrary procedure which has relied upon the seeming similarity of topography or function, rather than upon any empirical base. Although category construction will always be partially arbitrary, the possibility of forming response classes in terms of some type of

⁴¹Lytton, pp. 651-684.

⁴²Mash, Terdal and Anderson, "The Response Class Matrix," pp. 163-164; Patterson *et al.*, "Manual for Coding Family Interactions."

multivariate analyses⁴³ or analyses of stimulus control⁴⁴ has the potential for providing more than just nominal properties to behavior classes. For example, in an important study by Patterson and Bechtel,⁴⁵ using a multiple baseline design, members of an empirically defined response class were shown to vary together as a function of intervention, whereas a response empirically defined as a nonclass member did not so vary. This type of empirical category definition has both practical and theoretical implications. Knowing which response, for an individual, belongs to a particular class would provide direction in the implementation and monitoring of intervention programmes. In addition, determining the factors (chemical or environmental) which produce response classes, should provide impetus for understanding the mode of action of a particular treatment strategy.⁴⁶

Thirdly, a shift to dependent measures consisting of *units of interaction* in addition to single responses had a number of methodological implications. It resulted in the increased use of sequential coding systems which attempted to place responses in a temporal relation with respect to other individuals being observed.⁴⁷ Within a social learning framework, such systems, with their time-series data, provided a basis for determining the antecedents and consequents for a given behavior, the prerequisite for the establishment of functional relationships. They also provided information for developing a behavioral diagnosis,⁴⁸ and subsequent planning of treatment approaches. While the pay-offs are high with sequential coding systems, the trade-offs may be substantial. Methodologically, relative to nonsequential systems they require a great amount of time for observer training, accuracy, agreement, and generally a much more sophisticated and complicated system for data analysis.

Sequential coding systems, in order to maintain the temporal properties associated with behavioral occurrences, are frequently designed to be exhaustive. In this context, exhaustiveness refers to the property of systems which allow for the classification of any behavior with an existing code, such that there is never a no-score or nonoccurrence.⁴⁹ This property is consistent with

⁴³H. C. Quay, "Patterns of Aggression, Withdrawal and Immaturity," in H. C. Quay and J. S. Werry (eds.), *Psychopathological Disorders of Childhood* (New York: John Wiley, 1972).

⁴⁴Patterson and Cobb, "A Dyadic Analysis," pp. 72-129; G. R. Patterson and B. Bechtel, "Formulating the Situational Environment in Relation to States and Traits," in R. B. Cattell (ed.), *Handbook of Modern Personality Study* (Chicago: Aldine, 1971).

⁴⁵*Ibid.*, p. 348.

⁴⁶Patterson and Cobb, "A Dyadic Analysis," pp. 72-129; G. R. Patterson, "Changes in Status of Family Members as Controlling Stimuli: A Basis for Describing Treatment Process," in *Behavior Change: Methodology, Concepts, and Practice*, ed. L. A. Hamerlynck; L. C. Handy, and E. J. Mash (Champaign, Ill.: Research Press, 1973) pp. 169-191.

⁴⁷E. J. Mash and J. D. McElwee, "Manual for Coding Family Interaction," in *Behavior Therapy Assessment: Diagnosis, Design and Evaluation*, ed. E. J. Mash and L. G. Terdal (New York: Springer Publishing Company, Inc., 1976); Mash, Terdal, and Anderson, "The Response Class Matrix," pp. 163-164; Patterson *et al.*, "Manual for Coding Family Interactions."

⁴⁸F. H. Kanfer and G. Saslow, "Behavioral Diagnosis," in *Behavior Therapy: Appraisal and Status*, ed. C. M. Franks (New York: McGraw-Hill, 1969) pp. 417-444; M. R. Goldfried and R. N. Kent, "Traditional Versus Behavioral Personality Assessment: A Comparison of Methodological and Theoretical Assumptions," *Psychological Bulletin*, 77 (1972), 409-420; M. R. Goldfried and J. Sprafkin, *Behavioral Personality Assessment* (New Jersey: General Learning Press, 1974).

⁴⁹H. Guetzkow, "Unitizing and Categorizing Problems in Coding Qualitative Data," *Journal of Clinical Psychology*, 6 (1950), 47-58; R. W. Heyns and R. Lippitt, "Systematic Observation Techniques," *Handbook of Social Psychology*, Vol. 1, ed. G. Lindzey (Reading, Mass.: Addison-Wesley, 1954), pp. 370-404; Mash and McElwee, "Situational Effects on Observer Accuracy," pp. 367-377.

behavioral assessment in that it allows for the description of antecedent and consequent events. If this information were missing, as might be the case with a nonexhaustive scheme, it would not be possible to do a complete functional analysis. However, in terms of data analysis, exhaustiveness contributes to ipsativity in behavioral observation scores. If there is a greater amount of one behavior, there must necessarily be a lesser amount of another, and vice versa. Jones considers some of the problems associated with the use of ipsatized measures in data analysis.⁵⁰

So far, in tracing developments within the case-study phase of behavior modification, we have seen several changes in subject and treatment focus creating methodological problems associated with changes in the measuring instruments employed. Specifically an initial transfer of the operant paradigm from animal laboratory to human intervention has necessitated changes in the measuring instruments, as did the shift from an "operant set" in behavior modification to more of a social learning orientation. What is left towards the end of the case study stage, is a fantastically heterogeneous grouping of measurement instruments used by behavior modifiers.⁵¹ As they have evolved, these instruments have in common a general emphasis on naturalistic observation, complexity in terms of numbers and qualities of behaviors described, descriptions of units of interaction for determining stimulus-response-consequent relationships, and utilization of behavior categories. However, in spite of these commonalities, standardized and widely used measurement tools are almost nonexistent. This contention is supported in a recent report by Kanfer, in which the lack of uniformity in behavioral assessment is pointed out.⁵² In concluding this report Kanfer makes several recommendations:

- (1) Establish a pool of procedures now used to evaluate behavior change and construct from them a set of general procedural rules which can be applied to individual cases. This pool would simply be a list of practical methods for the observation of behavior both prior and during treatment.
- (2) Work toward increased uniformity and structure in the "informal" diagnostic procedures, such as interviews, behavior sampling, diary-type charts. The broad objective would be to permit research with comparable techniques and to develop approaches that yield information that transcends the limits of the sample parameters.
- (3) Isolate and refine instruments aiming at assessing different elements of the problem situation, e.g., the biological repertoire, the situational determinants, the locus of discomforts of conflicts, and the limitations of the therapeutic environment. Among these, measures of the demand characteristics and the available reinforcers in particular social environments are of critical importance, but almost nonexistent.
- (4) Work toward similar requirements of reliability and validity as is currently required to "standardized" tests for such individualized procedures as behavioral observation or self-observation.
- (5) Work toward understanding the different social and personal criteria for "improvement" so that inter-individual comparisons can be made, even though the clinician must adjust the techniques to suit the particular setting and client.⁵³

COMPARATIVE AND FOLLOW-UP STUDIES

These recommendations regarding measuring instruments and behavioral assessment are especially pertinent for behavior modification as it moves into

⁵⁰R. R. Jones, "Intra-individual Stability of Behavioral Observations: Implications for Evaluating Behavior Modification Treatment Programs," paper presented at the Western Psychological Association Meetings, Portland, Oregon, April, 1972.

⁵¹F. H. Kanfer, "Assessment for Behavior Modification," paper presented at a symposium on Newer Approaches to Personality Assessment, to the American Psychological Association Convention, Washington, D.C., September, 1971.

⁵²*Ibid.*, p. 1.

⁵³*Ibid.*, pp. 11-12.

the comparative and follow-up stages. For example, the establishment of a "set of general procedural rules" for selecting the appropriate measurement technique could come to include the factor of "efficiency" as described by Patterson.⁵⁴ Working "towards increased uniformity and structure" certainly would be a prerequisite for comparative studies.⁵⁵ The recommendation for "reliability and validity" would also facilitate comparative study.⁵⁶ The last recommendation involving working toward understanding different criteria for "improvement" is germane at several levels. First, as Kanfer states, it allows the making of inter-individual comparisons. Second, as discussed by some authors,⁵⁷ it allows for the making of comparisons between different modes of treatment. By referring to "different criteria," Kanfer brings up what is believed will be characteristic of measuring instruments in the emerging comparative and follow-up stages for behavior modification. This will be the use of multiple outcome criteria (measures) with a concern for "convergent validity."⁵⁸ If there are to be comparisons of different treatment strategies, which are based upon differing theoretical orientations it will be necessary to use a variety of measures. For example, treatment approaches concerned with more affective/cognitive questions might wish to employ questionnaire-verbal report data, psychopharmacological approaches might, perhaps, wish to obtain more direct physiological measures of outcome. Neither of these approaches would preclude the collection of concomitant behavioral measures, nor would a behaviorally based treatment strategy preclude the collection of more subjective or physiological measures.

So far, behavior modification studies reporting follow-up data have been minimal. Even where follow-up data is reported, it has tended to be only for a short period of time and in an unsystematic form. The appearance of studies, over the next decade, which attempt to assess the long-term effectiveness of a variety of treatment approaches, should illuminate a number of the methodological concerns expressed in this section. Clearly the question of follow-up will interact with the multiple outcome criteria approach. In order to talk about effectiveness over time, some specification of what constitutes "improved" must be indicated. Follow-up may affect different measures differentially. For example, questionnaire data may remain stable over time, while behavioral observations may change drastically, or vice versa.

It is likely that in attempting to do follow-up studies behavior modification researchers will encounter some of their most interesting and challenging methodological and conceptual problems. One reason for this is that in many ways the demand for maintenance over time contradicts some of the stated formulations of social learning theory.⁵⁹ The situational changes which are likely to take place over months or years should result in some degree of variability in behavior, over and above what would be expected in a relatively constant stimulus environment. How, then, is one to know whether

⁵⁴Patterson, "Intervention in Homes of Predelinquent Boys," p. 1.

⁵⁵O'Leary and Kent, pp. 69-96.

⁵⁶Johnson and Bolstad, "Methodological Issues," pp. 7-67.

⁵⁷O'Leary and Kent, pp. 69-96.

⁵⁸Johnson and Bolstad, "Methodological Issues," pp. 7-67; M. R. Yarrow, "Problems of Methods in Parent-Child Research," *Child Development*, 34 (1963), 215-226.

⁵⁹W. Mischel, *Personality and Assessment* (New York: John Wiley & Sons, 1968); W. Mischel, "Toward a Cognitive Social Learning Reconceptualization of Personality," *Psychological Review*, 80 (1973), 252-283.

maintenance or nonmaintenance of treatment effects is a function of the longevity of a particular treatment or of a change in the situational determinants perhaps interacting with some ongoing process? In fact, the high probability of a changing post-treatment environment has led some authors⁶⁰ to view follow-up data as relatively unimportant. However, others attempt to deal with the question of the reliability of behavioral measures over time.⁶¹ These authors are of the opinion that some degree of reliability (consistency over time) must obtain if meaningful statements about behavioral measures are to be made. It would seem that in order to obtain reliability over time, some type of redefinition of the "response measure" is required. This redefinition could take three potential forms.

First, it is more likely that a global or categorically defined dependent measure will exhibit greater reliability. So, for example, many of the measures reported which represent summations over a variety of specific behaviors, have been shown to be reliable, in spite of the fact that there may be variability of the particular responses making up the measure.⁶² A global "deviant behavior score" may be reliable, though, over time, the individual behavioral components making up that score may not be. The use of categorical measurement is seen as a desirable, if not necessary, condition for making the kinds of comparisons needed for long-term follow-up. Suppose, for example, a therapist was to treat Max (age two) and reduce his rate of defecation, food smearing and drooling. If these specific responses represented three dependent measures and a ten-year follow-up was done, it would be unlikely that Max (now age twelve) would be engaging in these behaviors. It would be absurd to conclude that treatment had long-term effectiveness. However, if a no-treatment control was used and it was found that ten years later this group also showed none of these behaviors, it would be equally absurd to conclude that our treatment did not have long-term effectiveness. Clearly, there is an erosion of the dependent measure over time — the responses used to measure effectiveness at one point in time, may no longer be appropriate at a later point in time. If, however, categorical assessment is used, it is possible to come up with a common measuring stick for seemingly different responses. Thus, both food smearing at age two and profanity at age twelve might both be defined as "deviant" — and comparisons (as would be done in follow-up studies) can be made. Empirically defined response classes (as previously described) might be developed. However, thus far, behavioral assessment procedures have used primarily arbitrarily defined response classes.

The rudiments for a second possible type of response redefinition, which would alleviate some of the difficulties for both comparative and follow-up studies with respect to the reliability of the dependent measures, are presented by Jones.⁶³ Regarding reliability for time series data, he presents the idea of conceptualizing reliability "in terms of the subject's behavioral predictability." For sequential observational data, this reconceptualization could involve the use of measurement terms which reflect the *structural* properties of an

⁶⁰Bijou and Peterson, pp. 63-78.

⁶¹Johnson and Bolstad, "Methodological Issues," pp. 7-67; Jones, "Behavioral Observation Frequency Data," pp. 119-145.

⁶²Johnson and Bolstad, pp. 7-67.

⁶³Jones, "Intra-Individual Stability," pp. 4-5; Jones, "Behavioral Observation Frequency Data," pp. 119-145.

interaction. With this formulation, as with the categorical redefinition described above, the specific responses in an interaction may vary, while the structural components show consistency over time. So, for example, the conditional probability that a teacher responds to a particular child in a certain way may be .90. While the frequency of responses of the child or teacher may vary over time, the probabilistic relationship between these two teacher-child responses may still be quite reliable.

Using the structural properties of interactional data as a response measure becomes even more evident if one describes the interaction in terms of a "redundancy" measure similar to those used in the information processing literature.⁶⁴ Redundancy with respect to the informational content within a particular interaction would also represent a measure of predictability, and has the potential for being independent of the specific content in that interaction. Some preliminary work attempting to relate this structural property of interaction to observer accuracy suggests that observing redundant interactions may adversely affect observer accuracy.⁶⁵

A third, and probably most difficult, reformulation of reliability for time series data would be in terms of the *functional* properties of the interaction. At this level, it would be possible to establish empirically the reinforcing properties of a particular behavior, in relation to one or more than one of the responses it follows. If the response is a generalized reinforcer, then again, variability over time for specific behaviors could be obtained, while the reinforcing properties of certain behaviors remain the same. Functional properties of an interaction may also be defined in terms of antecedent stimulus control. A procedure for this kind of formulation is presented by Patterson.⁶⁶ Patterson has developed an ingenious conceptualization and methodology for examining stimulus control properties both in terms of intra-individual response comparisons, as well as, comparisons across different family members. He has also tentatively demonstrated that this measure has "some degree" of sensitivity in reflecting changes which accrue as a function of intervention procedures. The unique aspect of using response measures which are defined in terms of the functional properties of interaction, whether in terms of reinforcement or stimulus control, is the possibility of obtaining reliable measures in spite of the fact that there may be considerable variability over time with respect to frequency and/or rate of occurrence of specific responses.

Another point to be made with respect to changes in measuring instruments concerns whether the observations obtained are to be used for the setting up of intervention programmes, for the evaluation of their outcome, or for both. Assessment procedures which are to be used for the setting up of individualized treatment programmes are likely to be more idiosyncratic than those used for the evaluation of treatment outcome. It may be possible to make broad statements about the effectiveness of treatment based upon much more global measures and behavior category definitions than would be necessary to make statements about the individual. As there is an increase in comparative studies

⁶⁴F. Attneave, *Applications of Information Theory to Psychology: A Summary of Basic Concepts, Methods, and Results* (New York: Holt, Rinehart & Winston, 1959).

⁶⁵E. J. Mash and G. Makohoniuk, "The Effects of Prior Information and Behavioral Predictability on Observer Accuracy," *Child Development*, 46 (1975), 513-519; Mash and McElwee, "Situational Effects on Observer Accuracy," pp. 367-377.

⁶⁶Patterson, "Changes in Status of Family Members," pp. 169-191.

it is believed that global outcome measures will be sufficient to answer the actuarial question as to which type of programme may be most "effective for a given population." The contributing factor here is that of cost, in that the use of global outcome measures is likely to be much less expensive. It is believed, however, that as more studies are designed to deal with the mode for a particular treatment, it will again be necessary to revert back to highly specific measuring tools. In summary, the transition of behavioral treatment through the case-study stage into a comparative and follow-up stage as described by Bachrach and Quigley⁶⁷ has resulted in a number of changes in the types of measuring instruments employed, which have in turn presented researchers with a series of related methodological problems.

In the same way that the development of behavior modification as a treatment strategy seems to have resulted in changes in the measuring instruments employed, there also seems to be a changing emphasis with respect to the types of design strategies used. The early emphasis upon single subject designs⁶⁸ while appropriate to the types of questions being asked during the case-study period, may be less appropriate for behavior modification as it moves into the comparative and follow-up stages.⁶⁹ The comparative stage could in fact be defined in terms of an increasing use of group designs. The critical issue of design is control, and the group design with adequate control groups should circumvent some of the problems involved with using reversals or multiple baseline controls.⁷⁰

A movement towards the use of group designs will also necessitate the use of a larger subject population. Undoubtedly, this will interact with the type of measures used (as previously discussed), in that highly complex, highly idiosyncratic measures are likely to be too costly for use. Given that larger subject populations are to be used, it will also be necessary to insure that the subject pools represent a fairly homogeneous grouping of individuals, with respect to the type of behavioral deficits or excesses exhibited.⁷¹ To the extent that homogeneous groupings can be used, it will be possible to reduce within group variability, and consequently, increase the likelihood that statements about the group will apply to the individual.

The use of group designs is likely to involve both comparisons across different types of treatment approaches (including no treatment controls), as well as comparisons between different types of behaviorally oriented treatment techniques. These two strategies are likely to yield different informational output. Comparisons with other treatment approaches address themselves to both social and theoretical questions, but it is believed that they are motivated more by societal needs and/or pressures.

⁶⁷Bachrach and Quigley, pp. 482-560.

⁶⁸M. Sidman, *Tactics of Scientific Research* (New York: Basic Books, 1960).

⁶⁹O'Leary and Kent, pp. 69-96.

⁷⁰Hall *et al.*, "Teachers and Parents as Researchers Using Multiple Baseline Designs," *Journal of Applied Behavior Analysis*, 3 (1970), 247-255; D. P. Hartmann and C. Atkinson, "Having Your Cake and Eating it Too: A Note on Some Apparent Contradictions Between Therapeutic Achievements and Design Requirements in N=1 Studies," *Behavior Therapy*, 4 (1973), 589-591; Kazdin, pp. 517-531; H. Leitenberg, "The Use of Single-Case Methodology in Psychotherapy Research," *Journal of Abnormal Psychology*, 82 (1973), 87-101.

⁷¹Kanfer and Saslow, pp. 417-444.

For example, regarding the control group design, Donald Baer makes the following point:

This is not to suggest that the group design is the proper replacement for single subject designs in future research. The questions which these control groups are to answer are actuarial questions, not analytical ones. They do not ask how behavior works; they ask only if it will usually solve the referring problem to get these behaviors arranged in a new way. We will no doubt continue to analyze behavior in the single case, for all the excellent reasons which Sidman made clear over a decade ago (1960), most of them bound to the thesis that it is in the single case that behavior resides, and so it is in the single case where it can best be analyzed. Our new problem is similar to that of any insurance company: we merely want to know how often a behavioral analysis changes the relevant behavior of society toward the behavior, just as the insurance company needs to know how often age predicts death rates.⁷²

It is believed that group designs may be misguided when they attempt to convince people with other viewpoints that "one way is best." The likelihood of convincing others, is believed to be minimal. Comparisons across different theoretical approaches are only feasible if each of the different approaches is willing to accept some common type of outcome criterion. If such common outcome criteria become available, then comparative studies of this type will be feasible. If efforts are not extended into developing common outcome criteria, and these studies are conducted unilaterally (from any singular orientation) then it is believed that they will have minimal payoff. On the other hand, comparative studies which seek to compare differing treatment techniques within a consistent theoretical overlay, are likely to have greater payoff. It is believed that this type of study will contribute to the refinement and efficiency of behaviorally based treatment techniques. This sort of comparative study, if it produced significant interactions, has the potential for describing which types of interventions are most appropriate to particular populations.

Another concern for behavior modification follow-up studies is with how long the period of follow-up should be. As previously indicated, the length of time of follow-up interacts with properties of the measuring instruments, as well as a changing stimulus environment. A key question is whether or not it is possible, on an a priori basis, to determine what constitutes an appropriate period for follow-up. Thus far, where follow-ups have been conducted, the time at which follow-up data were collected has been arbitrarily arrived at. Often, follow-up data are collected at a time dictated by the practical exigencies in the situation, rather than as a function of some pre-determined rationale. The fact that two weeks seems like a short period of time whereas two years seems like a long period of time has been used to suggest that two years is a better follow-up period than two weeks. This may not necessarily be the case. It is believed that the length of time necessary for follow-up is likely to interact with the particular behaviors which are under question, especially in situations where there is a progressive longitudinal erosion of our dependent measure. For some behaviors it may be possible that after two weeks, the level of a particular behavior is predictive of its level two weeks subsequent. If this does in fact turn out to be the case (as may be empirically determined) then it may be feasible to just collect two-week follow-up data. In the initial stages, however, it will probably be necessary to collect frequent follow-up observations. Clearly, the length of time for follow-up must also

⁷²"Behavior Modification: You Shouldn't," in E. A. Ramp and B. L. Hopkins (eds.), *A New Direction for Education: Behavior Analysis 1971* (University of Kansas: Department of Human Development, 1971).

be considered in terms of the rate of a behavior during baseline. If a particular behavior is a low frequency event, and occurs with a base rate frequency of two times a year, then a six-month follow-up which showed no behavioral occurrences, would give us very little information with respect to the long-range effects of our treatment. On the other hand, if a different behavior had a high base rate frequency of occurrence, for example, if it occurred ten times a minute, then a follow-up period of two days in which the behavior had a zero rate of occurrence might prove to be sufficient in terms of predicting the long-term effects of treatment.

The increasing appearance of comparative and follow-up studies should also necessitate a greater emphasis on the collection of normative data, as well as an increasing emphasis on longitudinal studies. If comparisons are to be made, there should be some estimates of base rates for a variety of deviant and nondeviant responses collected across time, across situations, and across populations. The methods for modifying "deviant" response repertoires are difficult to evaluate, if we are unable to describe what the "normal" or at least "average" response is for a given population, in a given situation, at a particular point in time. A number of recent studies⁷³ have begun to concern themselves with this question, and it is likely that there will be an accretion of this type of normative study through the next decade.

SUMMARY AND CONCLUSION

This paper has attempted to relate treatment and methodology for behavior modification by tracing progressive developments through the case study, and beginning comparative and follow-up stages of treatment, and pointing out how these developments necessitated changes in the types of measuring instruments employed including concern for both design and data analysis. It is believed that this concomitant and interactive development of treatment and method serves as a basis for understanding some of the current methodological problems for behavior modification. Included here are questions relating to the use of human observers in naturalistic settings (observer effect, observer bias, etc.), the reliability, validity, and generalizability of observational data, issues of experimental design and control, and finally concerns with respect to data analysis. In addition, it is possible that an historical perspective on the mutual development of treatment and method might provide a basis for understanding and/or predicting the directions that future developments might take. It seems likely that the intervention strategies developed (and the types of problems behavior modification addresses itself to) over the next decade, will precede, and probably be predictive of the methodological challenges which

⁷³M. E. Bernal *et al.*, "Comparison of Boys' Behaviors in Homes and Classrooms," paper presented at the Banff International Conference on Behavior Modification, Banff, Alberta, Canada, 1974; L. F. Delfini, M. E. Bernal, and P. M. Rosen, "Comparison of Normal and Deviant Boys in Their Homes," paper presented at the Banff International Conference on Behavior Modification, Banff, Alberta, Canada, 1974; S. M. Johnson *et al.*, "How Deviant is the Normal Child? A Behavior Analysis of the Preschool Child on His Family," paper presented at the Fifth Annual Meeting of the Association of the Advancement of Behavior Therapy, Washington, D.C., 1971; L. Terdal, R. H. Jackson, and A. M. Garner, "Mother-Child Interactions: A Comparison Between Normal and Developmentally Delayed Groups," paper presented at the Banff International Conference on Behavior Modification, Banff, Alberta, Canada, 1974; H. M. Walker, R. H. Mattson, and N. K. Buckley, "The Functional Analysis of Behavior Within an Experimental Class Setting," in *An Empirical Basis for Change in Education*, ed. W. Becker (Chicago, Ill.: Science Research Associates, 1971).

will ensue. Historically, changes in treatment emphasis, both from a practical and theoretical base, seem to have preceded development in methodology, as well as the concern for problems arising concomitantly as a function of these developments.

RESUME

Plusieurs des fréquents problèmes méthodologiques concernant la recherche et la pratique de la modification du comportement, sont considérés comme étant une conséquence directe des récents développements de ce nouveau sujet, en tant qu'effort pédagogique et thérapeutique. De façon à mettre ces problèmes en relief, les stratégies d'intervention du comportement sont étudiées tout au long des différentes phases du cycle de vie comme une méthode de traitement présentée par Bachrach et Quigley (1966): étude de cas, étude comparative, et étude de l'évolution du développement d'un cas. Dans chacune de ces phases, le type de questions du traitement posé, semble être un des facteurs primordiaux du genre de rapports méthodologiques qui se sont développés. Ces rapports méthodologiques comprennent ceux qui sont rattachés à la sélection d'instruments de mesure, de stratégies données et des implications pour les analyses données.