Software Assist for Education and Social Science Settings: Behavior Evaluation Strategies and Taxonomies (BEST) and Accompanying Qualitative Applications

Tom Sharpe

Purdue University

John Koperwas

Educational Consulting, Inc.

For all correspondence contact: Tom Sharpe, Ed.D., 207 LAMB, Dept. of HKLS, Purdue University, West Lafayette, IN 47907; phone: 765-494-0048; e-mail: tsharpe@sla.purdue.edu.

To the observational setting that
has blessed me with the most insight—
my immediate and extended family.

Table of Contents

---

for information:

SCOLARI
SAGE PUBLICATIONS SOFTWARE

Sage Publications, Inc.
2455 Teller Road
Thousand Oaks, California 91320
order@scolari.com
805.499.1325

Sage Publications Ltd.
6 Bonhill Street
London EC2A 4PU
United Kingdom

Sage Publications India Pvt. Ltd.
M-32 Market
Greater Kailash I
New Delhi 110 048 India

Acquiring Editor: Jim Gilden
Editorial Assistant: Sherry Miller
Production Editor: Jason Love
Typesetter/Designer: Rose Tylak

Author Acknowledgments

Tom Sharpe is an associate professor and teaching/coaching education program director in the Department of Health, Kinesiology, and Leisure Studies at Purdue University. Tom draws from a wealth of K-12 and postsecondary teaching and coaching experiences and many years of education and social science work at Research I institutions. Beginning with training from one of the leading applied behavior analysis psychology programs in the country, Tom's work lies largely in the education and social science application of sequential behavior analysis methods and in related computer-based tool development. He has authored more than 75 refereed articles and book chapters and is a regular contributor to the principles and practice of behavior analysis in teacher education and education research.

John Koperwas has been a practicing computer programmer for the past decade who, through the original support of S&K Computer Products, Ltd. and NEC Computers, pioneered the development of one of the first technologically based observational instruments widely adopted by teacher educators, clinical and experimental psychologists, and special educators throughout higher education. After a prosperous research career with S&K Computer Products, Ltd., John went into research and development partnership with Tom Sharpe and founded Educational Consulting, Inc., which now jointly serves clients worldwide ranging from teacher education programs to medical rehabilitation clinics to public school districts to special education and activity-based outreach centers all interested in the continuing development and use of the computer tools and related information contained in this text.

Preface

As computer hardware development moves forward with increasing speed and sophistication, our potential for as yet unrealized description and analysis of the world around us is greatly enhanced. Those of us engaged in the development of increasingly capable observational instruments are continually amazed at not only the wealth of additional information that becomes available but also the opportunity afforded to look at things in new and different ways. In these regards, these materials include a variety of behavioral, quantitative, and qualitative tools for collecting and analyzing direct observations.

We hope users will agree that this manual and set of computer-based tools are well designed for just these purposes. First, we wanted to provide researchers, educators, and clinicians multifaceted means to more capable and inclusive description of highly complex and interactive settings. Second, we wanted to provide tools not only capable of such a task but also capable of being programmed for the unique observational interests of a host of professionals working within a wide variety of situations and settings. Third, we wanted to draw on the conceptual work of pioneers in observational and field theory to provide a set of alternative lenses with which to view and analyze the rich and varied information that state-of-the-art computer-based data collection tools are capable of providing. And last, we wanted to include conceptual and procedural information (refer to pdf Chapter files), which may be helpful to those interested in the direct observation of complex phenomena as they undertake the challenges of such an endeavor.

In this last regard, we would like to thank the many professions and the dedicated scholars within those professions who have guided our thinking in this area and guided our tool development through its infancy. Academic programs too numerous to mention, from teacher education to special education to medical rehabilitation to applied and experimental psychology to ethology, have contributed in substantive ways to the form and character of the materials we provide. We could not have developed these materials without the guidance of those who have relied on earlier versions of this software and who have provided us with continuous advice and insight as they put our tools to use.

The software and related information contained within is designed for a wide range of professionals who have a common interest in the direct observation of complex and highly interactive phenomena. Based on earlier IBM DOS-based versions originally developed through Educational Consulting, Inc., the software is intended to accommodate a wide range of observational uses, with software programmers on staff available to design and develop additional applications based on the unique needs of various users. We hope that by providing a multilens approach to the direct observation and analysis of events, we will help users to gain greater insight into their respective settings of interest.

Chapter 1: The Challenge of Counting Things

> All things prepare the event. Watch.
>
> T. S. Eliot (*Murder in the Cathedral*)

Though more traditional descriptive and statistically based approaches to counting events in quantitative ways have frequented the education and social sciences in past decades, qualitative approaches to description and analysis recently have risen in popularity across the various social science disciplines. The information and accompanying software tools contained within these materials are provided to offer a unique combination across quantitative and qualitative approaches for professionals searching for enhanced or alternative means of performing their research and evaluation tasks. Much of the information contained within, whether conceptual, theoretical, or empirical, can be found in more detail elsewhere and is therefore summarized (and redundantly so for some readers, and for that we apologize) here with citations provided for the reader interested in greater detail. At these materials' core is the provision of a sophisticated, but very user friendly, software tool the development of which has become feasible based on recent and rapid advances in computer technology. This tool is designed to facilitate the collection and analysis of quantitative data in descriptive and analytic ways and, in particular, the counting of multiple events as they rapidly occur and interact in various ways within the complex natural setting in which they reside. In this regard, what might first be helpful to users of the software and related theoretical orientation that guides its use is a brief summary of two predominant methods of doing science in education and social science settings. This chapter provides a brief (and to some, perhaps, a simplistic) summary in the context of how each method may compliment and contribute to the information provided by the other and how each method is contained within the software tools provided.

At risk of oversimplifying a from-the-outside-looking-in perspective, the basic science community oftentimes views many of the published forms of education and social science in applied settings with great skepticism. This skepticism can in some ways be articulated as a tension between “art” and what the basic science community views to be genuine “science.” Long a marginalized research community by mainstream basic scientists, education and social science researchers have wrestled with the challenge of how to accurately and inclusively describe applied settings in a quantitative way. This has been particularly the case with those settings in which a host of complex behavior-behavior and behavior-environment relationships are ongoing with high frequency and in simultaneous or overlapping fashion.

Two dominant systems of thought in the social sciences that have tried to meet this challenge, social constructivism and radical behaviorism, have clearly provoked a degree of resistance within the larger scientific community and have been intermittent subjects of attack (see Binder, 1994; Brown, 1980; Glaser & Strauss, 1967; Morris, 1984, 1992, for a more complete discussion of these issues). In addition, Newman’s (1992, p. 13) point ironically remains the case that “proponents of . . . these [two] viewpoints have rarely accepted the other as valid, and have been at philosophical war with one another” since their respective inception. We agree with Newman, however, that these two dominant methodologies contain more common characteristics than differences and perhaps could help with the often cited criticisms of the other.

Social Constructivism

Social constructivists typically propose a qualitative approach to the descriptive-analytic challenge described above. A rigorous set of scientific procedures has been developed to collect and analyze narrative descriptions of various natural settings. A similar set of procedures has also been developed to collect and analyze various types of interview responses designed to gain insight into participant perspectives and cognitions believed to be operating within those settings (see LeCompte & Preissle, 1993; Miles & Huberman, 1984, for a detailed discussion). This school of thought provides a potentially inclusive description of behavior-environment events and potentially provides data related to thinking and mind, getting at what many behavior analysts term private events or covert behavior. This perspective is also criticized, however, due to the procedure’s inherent subjectivity and lack of content-validated measures. In a worst case, this method has been castigated for providing vague and nonspecific information, with narrative data more closely resembling art than included as products of science.

It is not the purpose of this discussion to malign qualitative research efforts in the social sciences, however. Clearly, qualitative research efforts have provided an alternative lens with which to observe, describe, and analyze the complex and highly interactive set of behavior-environment events that characterize most education and social science settings of interest. Much information and new knowledge has been gained from rigorous research efforts using a qualitative framework. The point of this discussion is rather to point to the general challenge of description and analysis of applied social science settings and where we are as a research community in this regard.

Behavior Analysis

A dominant competing method, termed applied behavior analysis, is a branch of psychology that attempts to discover and apply the ways in which particular behaviors are acquired and maintained in particular settings (Cooper, Heron, & Heward, 1987). Akin to constructivist principles, a therapeutic criterion is also included within the methodology, which stipulates that treatments designed for behavior change must benefit the client in some productive way (Kazdin, 1982). Applied behavior analysts also have provided a rigorous set of procedures designed to describe and analyze settings of interest in behavioral ways. Based largely on the early work of Watson (1970) and culminating with contributions by B. F. Skinner (1989), and to an extent the conceptual writings of J. R. Kantor (1953, 1959, 1969), this methodology at its core proposes a behavior-behavior and behavior-environment stimulus→response relationship. Within this framework, various studies have shown that multiple stimuli may be linked if repeatedly presented together; that stimuli may successfully be substituted for one another to produce the use of desired behaviors and the extinction of undesirable behaviors; that new behaviors or responses can successfully be trained and maintained in the absence of that training within various natural settings; and that many behaviors may even be generalized to situations outside of the primary experimental setting under the proper experimental training conditions.

The body of work generated by applied behavior analysis methods is also founded on a disagreement with more subjective methods of information gathering and limits itself to the collection and analysis of observable events that have a definite beginning and end time and that are amenable to being counted in some way. Though methods of counting are many and range from simple number or frequency counts to percentages of experimental time, to rates, to rates of acceleration or deceleration, the primary focus of this method is that of counting observable phenomena in some quantitative way. This method of observational assessment has suffered as well, particularly from the perspective that it is too time consuming to be feasible in highly interactive settings and that as a result, the data collection and analysis process often excludes many relevant behavioral events or environmental stimuli that potentially affect the primary behaviors and events of interest. Observing and analyzing only a very limited set of behaviors or environmental events within an experimental setting using more traditional paper-and-pencil methods has also fueled the criticism that behavior analysis as a method tends to fragment a naturally occurring process in which isolated events are aggregated over time with presumptions being made about how those events may affect other events in some mechanistic or additive way (Doyle, 1990). Additionally, many readers of behavior analytic studies tend to make assumptive leaps with regard to an explicit (and to our mind wrongly assumed "causal") link between various stimuli and responses operating within a particular situation. In other words, collecting information on the relative frequency of a particular behavior or environmental event that is hypothesized to act in a stimulating way, and collecting the frequency of a response that has a hypothesized connection with the stimulus under observation, does not necessarily support the causal nature of such a connection. When these types of data are collected in complex environments within which multiple behaviors and setting events are continuously interacting, we feel that a more appropriate analysis involves stipulating a time-based conditional probability of a behavior or event's occurrence given the immediately preceding or succeeding presence or absence of another behavior or event of interest.

Three conceptual tasks therefore remain. First, we must stress the equal importance of counting things in relationship to other more subjective ways of information gathering when involved in a scientific or evaluation enterprise in the educational and social sciences. Next, and the focus of Chapter 2, we must provide an argument for the feasibility of a counting-oriented method of data collection and analysis in complex interactive settings, and we must provide a compatible scientific theory that explains in a general way the manner in which behavior-behavior and behavior-environment interactions occur and that supports the importance of counting things as a way of knowing.

Science as a Counting Enterprise

According to Kerlinger (1986), the term *science* refers to the act of describing the observable world around us in a systematic and controlled manner. If scientific activity is regarded as a descriptive enterprise, or at least is considered to be descriptive in part, then it is necessary to systematically and empirically ensure the accuracy of those descriptions and systematically and empirically test for changes in those descriptions based on the introduction of behaviors and events (i.e., treatments or interventions) into the descriptive mix. What is ruled out according to this sort of definition of science is the explanation of relationships among observable phenomena using metaphysical or unsubstantiated beliefs. Stated by Peirce (in Buchler, 1955), the scientific method satisfies all doubts or competing answers by collecting information on which our thoughts, beliefs, perceptions, biases, values, attitudes, and emotions have no effect on the data gathering process.

In other words, a scientist endeavors to ensure the element of objectivity in his or her collection and analysis efforts. In this way, it does not matter who performs the information collection or data analysis procedures, for if those procedures are followed correctly and with rigor, the information and conclusions will be the same regardless of the particular procedural user. In this regard, we agree with Kerlinger's proposition that while there exist a number of methods (i.e., various quantitative, qualitative, and behavior analytic procedures) for collecting and analyzing information scientifically, there is only one scientific approach.

If we are to base a set of recommended data collection and analysis tools on the above definition of science, and in particular to be used within the education and social science enterprise, then these tools should have the capability to objectively and inclusively collect and analyze observable phenomena using traditional quantitative methods—by counting the occurrence and duration of various events. In addition, these tools should have the capability of describing elements of bias due to various contextual factors not amenable to more traditional forms of quantification—by further describing those same events in qualitative ways.

Though potential drawbacks exist with various scientific methods, one important feature of a behavior analytic approach to science is the ability to count or quantify the salient characteristics of particular behaviors and events so that others may gain an accurate representation of the setting described. An important feature of qualitative methods or narrative descriptions, on the other hand, is the ability to describe and contextualize characteristics of behaviors and environmental events not amenable to counting. In the ideal, similar scientific goals are inherent to each method, and each method may complement the other for increased understanding if used in concert.

The goal of each observational method is to portray accurately the setting observed. One method attempts to quantify behaviors and events; the other endeavors to contextualize those events further with regard to the purposes of those events and the perceptions of participants within the setting observed. With each method, the natural setting is the direct source of data and the scientist the key information collection instrument. Each method endeavors to provide a complete and inclusive description of setting participants, the environmental setting, and the interactions that occur between participants and environment. Both methods are concerned with the interactive process rather than simply outcomes or products. While qualitative research is an inductive process that does not operate within any preconceived theoretical or hypothetical framework and prioritizes contextual meaning as the essential concern, and behavior analytic research typically operates within a predefined framework that guides the categorization and analysis of quantitative observations, we feel that each method operates in the ideal according to the same general scientific approach.

The primary focus of these materials is therefore to provide a more amenable and appealing means to the counting of things within the larger umbrella of the scientific enterprise. Though the computer-based tools provided also include a compatible qualitative feature, our efforts are due largely to the prevailing challenge voiced with regard to the difficulties of successfully gathering and analyzing direct observational data in quantitative ways. Our efforts are also guided by Gottman's (in Bakeman & Gottman, 1986) characterization of the social and educational sciences as using observational measurement of any sort being in the vast minority of current research practice. One way to change this characterization is to provide user friendly tools that both new and experienced members of the research and development community will be receptive to and that various professionals will see the advantages to using when held up against other ways of description and assessment. Gottman perhaps sums the importance of direct observational methods (whether qualitative or behavior analytic in form) best as follows:

> Observational methods deserve a special role in our measurement systems. First, we think that the descriptive stage of the scientific enterprise is extremely productive of research hypotheses, models, and theory. Second, the time is ripe for a reconsideration of observational techniques because we now know a lot more about what to observe, how to construct reliable measurement networks, and how to analyze data to detect interaction sequences. (p. xiv)

These materials place a primary emphasis on new and innovative ways of counting things and on some respective recommended analysis techniques when involved in a direct observation enterprise. It is the latter point that Gottman raises that we pair with facilitative computer technologies in the hope of providing an appealing observational tool and in the hope of stimulating an increasing number of professionals in the education and social sciences toward direct observation approaches to their chosen research and evaluation tasks.

Chapter 2: An Interbehavioral Field Systems Perspective

> The worth of a scientific system lies in its "usefulness and economy."
>
> B. F. Skinner

With regard to our introductory comments in the Chapter 1 pdf file, we clearly feel that the scientific act of counting things is not necessarily an approach that should be pushed aside as qualitative narrative methods become more predominant in education and social science activity. Grounded in behavior analytic theory, direct observational methods that have focused on quantitative counting approaches to behavior-behavior and behavior-environment interactions have, however, suffered from some important criticisms. These criticisms, in part at least, may have led education and social scientists to search for other methods that may provide a more capable observational lens. Even in the proverbial hey day of behavior analysis in education research, Rosenshine and Furst (1973) articulated a cautionary tale with regard to the descriptive counting of behaviors and setting events. They voiced uncertainty with respect to a descriptive counting method's ability to discriminate between relatively more or less effective professional practice within and across particular educational and therapeutic situations. Argument was that it may be possible that behavior-behavior and behavior-environment interactions may be so complex and idiosyncratic to particular situations that specific interactions may not be isolated and generalized to ensure more broadly based therapeutic ends.

This argument remains today. Doyle (1990), for example, argues that while quantitative knowledge about professional practices is important, what typically occurs is the counting of only a few behaviors in isolation and the inappropriate fragmenting of a larger interactive process into discrete elements that are presumed to affect target behaviors or events in a causal and additive way. To avoid rendering the study and evaluation of education and social science settings simplistically generic, and to ensure that behavior analytic approaches remain in the mainstream of applied science, a more sophisticated method of direct observational data collection and analysis is necessary that is capable of quantifying the repeated occurrence of multiple variables and their many interactions in setting-specific context.

Taking education research as an illustration, the focus of behavior analytic research has traditionally been on the behavior of a particular teacher with a single metric of interpretation, usually a student behavior dimension such as ALT (Academic Learning Time; Metzler, 1989). Although some research designs examine the behavior of more than one person (Kazdin, 1982), rarely have behavioral methods focused on the behavior-environment interactions among multiple individuals (Wampold, 1992). As such, we agree with Doyle (1990), Kantor (1969), Morris (1992), and others that four recommended research and evaluation trends in education and social science efforts are apparent and that these trends beg the need for a more capable data collection and analysis tool. First, behavior-behavior and behavior-environment relationships are a very complex set of phenomena that need to be more thoroughly and inclusively examined. Second, emphasis must be placed on ecological validity, or analysis of the various settings in which these interactions are relatively more or less productive. Third, analytic emphasis must be placed on not just the discrete characteristics of behavior and environment events (e.g., number or frequency, percentage of experimental time, rate, etc.), but a complete analysis must also include the form and character of the multiple stimuli and response *relationships* among multiple behaviors and events as they actually occur in time in particular situations. Fourth, emphasis must be placed on the discovery of behavior-behavior and behavior-environment relationships that are predictive of therapeutic ends in particular situations.

Behavioral Science as an *Inter*behavioral Field

Consistently stated by current and past experimental and applied researchers interested in various behavior-behavior and behavior-environment relationships, the aim of behavior analysis has been the prediction and control of behavior through the descriptive counting of observable events (Morris, 1992). Throughout the research history[1] of behavior analysis, however, a predominant emphasis has been on the demonstration of potentially controlling events on particular client behaviors within a particular experimental or training situation, typically in a mechanistic fashion (Delprato, 1992; Morris, 1991). In addition, the predominant focus of analysis has been on the characteristics of the behaviors or events of interest themselves (such as frequency or percentage of experimental time).

Though perhaps viewed as a subtle difference at first glance, a more and more frequently seen and potentially appealing and productive approach to the analysis of behavior is one that emphasizes the discovery of the multiple conditions that tend to control various behaviors and how one may go about productively altering those conditions toward therapeutic behavior-change-oriented ends. Additionally, and related to the emphasis that an interbehavioral or field perspective[2] on the analysis of behavior brings to the mix, is an analysis focus on the probable interactions in time-based sequence among behaviors and events and the strong emphasis on the situations and contexts in which those interactions take place.

Illustrations from Education

Some illustrations may prove helpful here. In an education setting, for example, a typical nonbehavioral investigative approach might be to ask a teacher a series of questions and from the responses gathered provide a summary or assign a score indicating which tasks in a classroom were perceived most effective or satisfying. This type of information would allow us to try to determine relatively effective or ineffective instructional practices, or relate job satisfaction with other dimensions of that teacher's professional practices, all in a qualitative or subjective way. This type of measure, however, does not tell us much about how that teacher definitively interacts with his or her students on a regular basis. One productive means to learning more about the variable of professional effectiveness might be, therefore, to describe and analyze behaviorally how that teacher relates to students as various teacher-student interactions actually take place.

Once a behavior analysis approach is decided on, the appeal of an interbehavioral or sequential approach in relationship to a more traditional and static means can be easily illustrated. When studying the same classroom teacher, for example, you could decide on a static measure such as the frequency of instructional, feedback, and interpersonal teacher behaviors, and the related percentage of class time students devoted to behaviors such as skill practice, organization, or waiting and listening. This approach yields important information regarding how teachers and students tend to use their time in a particular situation. Many of these behaviors also are positively or negatively correlated with long-range measures of learning and achievement, yielding an indicator of the general effectiveness of the setting under observation. Using this more traditional static approach, however, still does not make explicit the sequential or interactive connection between the particular things a teacher does in that situation and how a student responds to those teacher behaviors.

Knowing how often instructional and organizational practices take place, or knowing how much time a student devotes to skill practice, does not tell us much about how those teacher-student behaviors were sequenced in time, nor does it tell us much about the moment-to-moment instructional and social interactions that are facilitative or inhibitive of recommended student practices. This is not because the terms and definitions included in the observational categories are inappropriate to potentially producing this information, but because the way in which the data are recorded on these behaviors is not conducive to capturing behavior sequences. Sequential data provide an additional level of information about the educational setting we might observe and provide information that is not accessible through other means. A sequential data set could, for example, help answer questions like (a) What do students do after certain types of instruction, and did that type of instruction help improve student skill practice efforts? and (b) What are the characteristic ways in which students tend to respond to different types of interpersonal interactions? Answers to these types of sequentially based questions require a different type of observational lens and a different way of collecting and looking at observational data.

While a more demonstration-oriented (refer to Morris, 1992, for a detailed discussion of *demonstration* versus *understanding* as related to applied behavior analysis methods) or traditional behavior analysis can provide information related to how much instruction or teacher reinforcement tends to be present before a student behavior is modified, sequential data provide information related to the discovery of how different student behaviors tend to be present as a specific function of time-based proximity with particular teacher practices in particular educational situations.

The Utility of an Alternative Lens

Two articles (Sharpe, 1997a; Sharpe, Hawkins, & Lounsbery, 1998) provide important illustrations of the utility of describing and analyzing behavior-behavior interactions in educational settings using an interbehavioral or sequential lens. In these materials, it is argued that one person's behavior is both a response to another's past behavior and a stimulus for yet another's future behavior and should therefore be examined sequentially. For example, a student's off-task behavior may be a response to a denied request of the teacher as well as a stimulus to another student's withdrawal from the ongoing activity *and* a stimulus to teacher behaviors designed to curtail the off-task episode. Therefore, while two teachers operating in two very different settings may exhibit similar frequencies or class-time percentages of certain behaviors, they may exhibit very different teacher-student interactional patterns. It is the description and analysis of the sequential nature of the behavior among teacher and students that may allow discovery of some important therapeutic relationships in each teacher's setting.

In the simplest case, a sequential data collection lens answers the question of whether one behavior follows another behavior more often than would be expected by chance. For example, does a particular student's off-task behavior increase the probability of a particular teacher discipline practice, and, if so, what does that teacher do in sequence to curtail future incidents of off-task behavior?

To talk about this type of example, it is first important to introduce how this type of data might be collected. In addition to traditional measures such as frequency or percentage of experimental time, a measure of a start time and stop time is also recorded for each behavior occurrence in a data record. Adding this type of information provides a time-based sequence of behavior-environment occurrences. Sample data from an experienced-novice public school teacher comparison (Sharpe & Hawkins, 1992) help illustrate the utility of a sequential data lens. For the sake of illustration, if we suppose that observational focus is on only teacher instruction (event A) and student engagement in the subject matter (event B), then a representative sequential four-minute data segment for a novice teacher might look something like the following:

ABAABABBABBAAABABBABAAABBAAABB

If we further assume for the sake of illustration that each behavior occurrence was of the same duration, a simple frequency, or *non*sequential behavior analysis would yield that A occurred 16 times and B occurred 14 times. A sequential analysis of this same data segment shows (a) an unconditional probability of A to be $p(A) = 16/30 = .53$; (b) an unconditional probability of B to be $p(B) = 14/30 = .47$; and most important, as will soon be demonstrated, (c) a conditional probability of B given the occurrence of A immediately before B to be $p(B/A) = .56$. The intent of a sequential analysis here is to reduce the uncertainty of B's occurrence given the knowledge of the immediately preceding event in the interactional chain.

If we then compare an experienced teacher's sequential behavioral chain that includes the same behavioral events to the novice data and were emitted in a similar teaching situation, the importance of this type of information is well illustrated. Again, suppose that focus is only on the behaviors of teacher instruction (event A) and student engagement in the subject matter (event B). If this is the case, then the following chain segment may be recorded for the experienced teacher for the same representative four-minute time period (again, refer to Sharpe & Hawkins, 1992):

ABABAABABABABAABABBABABABABABA

Observing this experienced teacher's observational data by traditional behavior analysis methods, A again occurred 16 times and B occurred

14 times. Stopping at this point would yield no differences in novice and experienced teacher behavior. Viewing the unconditional probability of A and B (p(A) = 16/30 = .53; p(B) = 14/30 = .47) also yields no differences across the experienced and the novice teacher data sequences. What is important to illustrate, however, is that analyzing the experienced teacher's data conditionally and sequentially provides a marked observed difference as follows: The conditional probability of the occurrence of B, given that A has occurred just prior to B, is the proportion of times that B occurs immediately after A—A occurs 16 times, and of those times, B occurs immediately after A 13 times. Thus, the conditional probability of B occurring dependent on A is p(B/A) = 13/16 = .84. Clearly, .84 is markedly different from the novice conditional and sequential probability of .56. Upon closer sequential analysis of the experienced versus novice teacher data this illustration draws from, potentially productive differences in the form and character of teacher observation and feedback in the context of the variable quality of student subject matter engagement, the proximity of the teacher in relationship to the student involved in the interaction, and other behavioral variables become evident.

While this is a very simplistic example, adding a sequential capability to the analysis of multiple occurrences of behavior-behavior and behavior-environment interactions as they naturally occur in time-based sequence allows quantitative description of the many different interaction patterns among different participants operating in various situations. In other words, to the degree that one interactive participant's behavior is dependent on the immediately preceding behaviors of another participant, that participant's behavioral response probabilities are altered as a function of the types of behaviors and environmental cues also operating in the setting under observation.

The Importance of Field Theory

An interbehavioral approach to the quantitative analysis of behavior is founded on the pioneering work of J. R. Kantor (1953, 1959, 1969). Much of Kantor's thinking, in turn, is founded on the more generic field systems theory originally postulated by Einstein and Infeld (1938). In agreement with contemporary methodological recommendations, Kantor felt that despite the many contributions of the experimental and applied analysis of behavior literatures, its lack of a field character left this body of work open to criticism. The inclusion of a field systems component to behavior analysis was, according to Kantor, an attempt to move forward from the strict lineal mechanics notion of a Skinnerian S→R relationship. To Kantor, the Skinnerian behavior-behavior and behavior-environment relationship model was lacking in that behavior and environment were described primarily in terms of the characteristics of the stimulus (S) and response (R) of interest, and causal assumptions were prone to be largely assumed and often placed inappropriately.

At risk of undue complexity, and lacking a practical methodology for carrying out his theories in applied research settings, Kantor (1969) stipulated that behavior-environment relationships should be viewed as a field, defined as

> the entire system of things and conditions operating in any event taken in its available totality. It is only the entire system of factors which will provide proper descriptive and explanatory materials for the handling of events. It is not the reacting organism alone which makes up the event but also the stimulating things and conditions, as well as the setting factors. (p. 371)

Stated in more familiar terms, Kantor posed that while study of the characteristics of individual behaviors to discover more about interactive relationships was important and productive, the most important dimension to interaction was that of the transactions or connections among multiple behaviors in setting-specific situations and not the isolated character of each behavior as a stand-alone entity. Though rooted in behavioral psychology, Kantor's work and the work of those who expanded on this premise were intended to reach a broad audience of professionals with interests in studying and evaluating human and other species at various stages of development and across many diverse settings. The connective thread that Kantor and these materials provide is a common interest in observing highly interactive social or educational behavior in applied settings. Primary focus is on which behaviors and events tend to precede or follow which other behaviors or events, and which of these sequential connections tend to be most productive in meeting particular educational or therapeutic ends.

A primary criticism of Kantor's contributions to the social and psychological sciences, however, is that what Kantor laid out as important to study stopped with the provision of a philosophy and did not include an applied methodology. In other words, though an argument in favor of both a more inclusive description of behavior-behavior and behavior-environment events and the importance of looking into the sequential nature of behavioral interactions was rigorously constructed, and many researchers, educators, and clinicians saw the potential importance of increasing the availability of this kind of information, an amenable data collection method that allowed the description and analysis of more inclusive and sequentially based information in applied settings remained largely unavailable.

Criticisms of Kantor's work aside, field thinking is an attempt to replace a more lineal mechanic scientific view. In the mechanical view, a single-headed arrow is typically used to show a one-direction cause and effect relationship between two events of interest and argues all other events as extraneous and hopefully under experimental control. The field concept, however, offers a double-headed arrow (see Morris, 1992, p. 15) between multiple events and aspires to describe probabilistic connections in time- and setting-based sequence.

Providing a tool that can (a) yield more inclusive quantitative descriptions of (b) multiple concurrent events that are (c) observed and recorded as they take place is what remains necessary to the application of an interbehavioral or field systems approach to education and social science concerns. As stipulated by Delprato (1992), such a tool must be capable of

> (a) identifying, defining, and measuring far more than the usual limited number of variables, and (b) tracking the [sequential] status of these variables on a moment-by-moment basis. (p. 3)

If such a tool were to be made available, then we feel that the direct appeal to education and social science concerns is readily apparent. Clearly, teaching, professional preparation, and most social interplay between human or animal participants is highly interactive and behavioral in character on many levels. These types of interactions involve what each participant is doing behaviorally with, and in response to, the others. These interactions also involve more subtle behaviors indicative of a qualitative or interpretive character related to emotional communication and attachment (Ray, 1992). Such interactions are conceivably quantifiable given the proper data collection and analysis tool, for we may now be speaking in terms of behaviors such as verbal intonation, rate of speaking and pausing, and nonverbal gestures, all challenging but approachable given the proper combination of behavioral and qualitative lenses with which to observe. Stated from the perspective of a hypercritical basic scientist, the art of teaching or social interaction does not have to remain the exclusive domain of artistic representation but is conceivably amenable to a more traditional and quantitative scientific description and analysis if (a) proper tools are provided for such a task and (b) education and social scientists are receptive to such an undertaking.

In these regards, a tool capable of collecting, describing, and analyzing a more inclusive volume and variety of behavior-environment data, and a tool that contains greater economy and flexibility of display, should prove helpful to researchers and evaluators alike whose subject matter is complex interaction in situational context. As further stipulated by Locke (1992), a quantitative/qualitative tool capable of putting into practice what Kantor recommends should provide data concerning “exactly what we are doing, as opposed to what we have come to imagine we are doing” (p. 86). For an elegant educational illustration of how such a tool might overcome the “illusion” that we know in intuitive ways how we or others have behaved in certain situations, refer again to Locke (1992, pp. 86-87).

The Implications of Computer Technology

Many education and social scientists currently hold the position with regard to their respective research and evaluation tasks that until data collection and analysis tools are developed that are perceived to be time efficient, cost effective, and easy to understand and use in comparison to other methods, those professionals will use those tools that are readily available and easiest to implement. When one compares the relative ease of implementing an instrument such as a questionnaire in contrast to the direct observational recording of behavior and event characteristics with a stopwatch and paper and pencil or some other traditional method, the choice for most professionals (and in particular, aspiring professionals such as graduate students) is an easy one. While there are other reasons to be sure, we believe like Bakeman and Gottman (1986), Binder (1994), and others that this characterized path of least resistance has acted in large part to diminish the use of quantitative behavioral methods in the education and social sciences.

What we are also firmly convinced of as well, however, is that the recent explosion of computer applications in our society has provided a strong catalyst for reconsidering the feasibility of direct observational research techniques because (a) we have learned from pioneers in the area of systematic observation and behavior analysis a lot more about what should and what could conceivably be observed (see Heward & Cooper, 1992; Morris, 1992; Skinner, 1968), (b) we have seen much more capable statistical measurement and reliability of measurement procedures developed (see Cohen, 1960; Gottman & Roy, 1990; Johnston & Pennypacker, 1980; Ray & Delprato, 1989), and (c) we now have access to alternative lenses to be used in the analysis of observational data, particularly with regard to interactive sequences (see Sharpe, 1997a, 1997b; Sharpe, Hawkins, & Lounsbery, 1998; Sharpe & Lounsbery, 1998).

Prior to the provision of sophisticated computer tools designed to collect and analyze multiple behaviors and events, and designed to allow the viewing of the sequential connections between those behaviors and events in interactive settings, researchers and professionals interested in studying and evaluating the relative effectiveness of those settings with a direct observation lens typically used traditional stopwatch and paper-and-pencil recording methods (refer to Kazdin, 1982, for a detailed description of accepted recording techniques). Though many education and social scientists voiced an interest in studying and evaluating multiple events in a more inclusive fashion and voiced an interest in the sequential aspects of interactive behavior, most settled for static observational measures due to the limited capability of the data collection and analysis tools available to them.

For example, a researcher interested in recording the number of feedback statements used by a youth sport coach and the related changes in the percentages of skill practice time by a representative athlete could be accommodated using a moment-by-moment recording procedure. In this case, the observer would simply break down the athletic practice time into logically specified time frames (usually one to two minutes) and spend one minute recording coach feedback occurrences, one minute recording the amount of time an athlete spent in skill practice, and one minute in data management tasks. Such a procedure did not allow, however, the recording of multiple coach and athlete behaviors and did not provide a data documentation lens that allowed observation of the connections among coach and athlete behaviors in time-based sequence. With the advent of computer-based data collection and analysis tool development, this need no longer be the case. In this light, computer tool development is providing the opportunity to look at interactive settings in more inclusive and in different ways, using different measures, and thereby illuminating the sequential nature of the many behavior-behavior and behavior-environment relationships in the setting under observation.

A Cautionary Note

If we are truly capable of enabling the type of description and analysis alluded to, then some potential concerns should be highlighted in relation to the development of computer applications (or any tools for that matter) designed under the guises of research and development, educational improvement, and clinical assessment. Clearly, the explosion of computer application development in the social and natural sciences is becoming a reality. However, our feverish cultural pursuit of new computer applications provides an inherent danger. We agree with Sharpe and Hawkins (1998) and Sharpe, Harper, and Brown (1998) that when intimately involved with computer application development, it is often tempting to move beyond developing an application to improve professional practice or extend a body of knowledge and develop that application for the pure sake of being able to do so. If we can manage to remain sensitive to this challenge, however, the potential for computer applications to provide tools that may uncover information previously unknown, and the potential for far more effective education and evaluation enterprises as a result, provides support for forging cautiously and systematically ahead with such development.

The primary question (posed by Sharpe & Hawkins, 1998, p. 20) related to computer application development is therefore one of “How can we avoid becoming technology’s servant, while at the same time explore how technology may better facilitate our educational ends?” Our culminating efforts within these materials to provide what we feel to be a comprehensive descriptive-analytic lens compatible with behavior analytic and qualitative frameworks is based on our past efforts and the efforts of many others who have pioneered a scientific trail before us in the areas of (a) the design and implementation of education and social science observational systems, (b) the development of educational and laboratory experiences for various preprofessionals in training, and (c) the exploration of some of the unknown determinants of effective practice across a host of education and social science settings. Researchers and clinicians involved in these types of activities have continually been interested in the development of data collection and analysis tools that might better serve their respective research, education, and clinical tasks, and it is upon a wide variety of their efforts that we have drawn. In other words, the computer-based tools that are provided with these materials have evolved based on a particular set of theoretical constructs and a particular set of data collection and analysis needs, and not simply because we have been enamored with our potential capability to do so.

In this regard, we define *technology* as more than the computer hardware, software, and electronic systems that are so often confused in our contemporary culture as synonymous with the end products of technology. Though necessary to the development of the type of scientific tools contained within, they are only the mechanical parts (much like an electron microscope is a mechanical part) that still require the trained professional to connect their use in appropriate ways to the discovery of new and important information and to enhancing educational, evaluative, and therapeutic applications. Relying again on the perspective of Sharpe and Hawkins (1998), the computer applications we are providing to the education and social science communities are merely tools and mechanical instruments much like the stopwatches and pencils were tools of the direct observer (and still are for some) before the advent of computer-aided data collection, or like mimeograph machines and chalkboards before the advent of advanced photocopiers and sophisticated computer-driven presentation software.

Related to our perspective of the scientific method, computer applications only gain value through their fruitful employment for those scientific purposes. One of the chief characteristics that some anthropologists have used to distinguish humans from other forms of life is their penchant for tool making. To the degree that this distinction is a legitimate one and to the degree that computer application development is a form of tool building, we argue that computer application development is a naturally evolving part of our endeavors to learn more about the world around us. Furthermore, if computer application development is endeavored according to the guiding principle just stated, then such efforts are an appealing facilitator of the scientific enterprise. The danger to which we must remain sensitive is one of not crossing the line between tool building for a redemptive purpose and tool worshipping (Ellul, 1964; Neitzche, 1978), which tends to dehumanize. It is in this light that we feel it important to caution the reader regarding the importance of steering clear of the primary enticement that the computer-based tool-building enterprise provides—that of developing and using a tool simply because we can and of becoming infatuated with these gadgets simply because they are interesting to our senses.

Perhaps Dwyer (1996), and relatedly Sharpe, Harper, and Brown (1998), have provided the best articulation of computer application development in relationship to scientific practice as follows: While technology may or may not add value to academic and professional practices in the long run, it will only realize any potential value if and when it becomes an integral part of a comprehensive and scientifically supported plan for professional improvement. Users of a particular computer application need to be prepared to both use it and document its various benefits and drawbacks in terms of one more potentially appropriate tool in their professional arsenal. We feel that a couple of introductory messages are clear based on this discussion. First, as a scientific community, we are far from realizing all of the possible methods of description and analysis that might be used to learn more about behavior-behavior and behavior-environment interactions contained in the world around us. Second, advancements in computer-based applications may afford a facilitating means to providing important alternative methods of data collection and analysis in the social sciences.

Endnotes

1. Refer to Kazdin (1982) for a detailed annotated discussion of the history and development of behavior analysis as a methodology.

2. Refer to Lichtenstein (1983) and Ray and Delprato (1989) for a detailed discussion of interbehavioral theory and practice and Einstein and Infeld (1938) for a generic discussion of field theory.

Chapter 3: Constructing an Observation System

> What each man knows is, in an important sense, dependent upon his own individual experience: He knows what he has seen and heard, what he has read and what he has been told, and also what, from these data, he has been able to infer.
>
> Bertrand Russell

Constructing a category system for direct observational purposes, describing the events contained in that category system in quantitative and qualitative ways, and applying mathematical models to analyze the discrete and sequential organization of event occurrences have frequented the education and social science literatures for the past three decades. In relation to sequential analyses, methods have included Markov chain analysis (Chatfield & Lemon, 1970), information theory (Bakeman, 1978), cross-spectral analysis (Gottman, 1979a), grammatical inference (Rodger & Rosebrugh, 1979), and the most frequently used sequential behavior analysis (Bakeman & Gottman, 1986; Gottman & Roy, 1990). It is this latter method first conceptually developed by Sackett (1979, 1980) in the infant development literature and formalized by Gottman's statistical modeling equations that we feel is one of the most appealing applications provided by the software tools included within these materials. A sequential approach to the description and analysis of observational data may also be an approach that is unfamiliar to many potential users of the observational software tools we provide. As such, it is important to devote space to a detailed illustration of this method's potential appeal to the study and evaluation of social science settings.

Application Illustrations

Sequential behavior methodology in particular has been applied with productive results to a variety of behavioral questions in the social sciences. Though this method remains unfamiliar to many researchers, a rich source of examples now exists in the literature. These illustrations should prove helpful to those interested in category system examples containing similar behaviors and events to those that a particular social scientist may wish to study. Illustrations of interest include examples in interactional rhythms (Scheflen, 1982), family therapy and marital interaction (Gottman, 1979b; Wahler & Hann, 1987), clinical psychology (Ruben & Delprato, 1987), school psychology (Martens & Witt, 1988a, 1988b), ecological psychology (Willems & Raush, 1980), health delivery services (Ray, 1983), general interpersonal skills (Faraone, 1983; Jacobson & Anderson, 1982), and communication ethology (Altmann, 1965; Mjrberg, 1972; Ray & Delprato, 1989). In addition, some of the work in microethnography (e.g., Erickson, 1982) in which short time frames are used to record in rich detail various subtleties of human interaction may be helpful.

Many category systems of a more traditional nonsequential nature also exist in the teacher education literature and may prove helpful as well. For readers interested in looking over examples of these types of category systems, we recommend reading Darst, Zakrajsek, and Mancini (1989) or Stallings, Needels, and Sparks (1987). The pioneering work of Flanders Interaction Analysis Categories (FIAC; Flanders, 1970) or the efforts of Dunkin and Biddle (1974), albeit with some easily recognized methodological concerns, may also prove helpful. Use of a more traditional category system approach to behavioral evaluation in educational settings is supported by a productive literature and includes the improvement of (a) general instructional practices (e.g., Ingham & Greer, 1992; Kamps, Leonard, Dugan, Boland, & Greenwood, 1991) and (b) specific competencies such as pupil attending, classroom control, and pupil praise (Cooper, Thomson, & Baer, 1970; Cossairt, Hall, & Hopkins, 1973; Hall, Panyon, Rabon, & Broden, 1968; Page, Iwata, & Reid, 1982).

The physical education (Darst et al., 1989) and special education (Miller, Harris, & Watanabe, 1991; O'Reilly & Renzaglia, 1994; Warger & Aldinger, 1984) professions are two teacher-training areas that have used direct observation approaches to the behavioral evaluation of teaching activities with great success when implemented by faculty supervisors. Most of the behavior analysis evaluation tools used for these purposes have been of a criterion-based nature in which teachers or teachers-to-be are held accountable to a target frequency, rate, or percentage of classtime use of known effective instructional practices and, in turn, the targeted behaviors of their pupils thought to be correlated with subject-matter acquisition are separately monitored (e.g., Carnine & Fink, 1978; Greer, 1985).

We argue that training individuals to criterion use of particular behaviors to be used in their respective professional settings falls under what Morris defines as "demonstration" (1992, p. 9), in which a rule-governed approach to applying theoretically effective behaviors is learned and consistently used by a client. An appealing alternative or addition to a demonstration approach is the addition of a sequential observation capability as offered by the computer-based tools contained with these materials. Providing a sequential analysis capability to data collection and analysis overcomes the challenge of identifying multiple stimulating events that set the occasion for desirable and not-so-desirable responses, all in situational context (Touchette, MacDonald, & Langer, 1985). If used appropriately, researchers and evaluators may facilitate what Morris terms "understanding" (p. 9), in which emphasis is placed on discovering just which behaviors tend to facilitate or impede the use of other behaviors within and across particular situations (see Sharpe & Lounsbery, 1998, for a detailed education example).

Although sequentially based contributions are relatively less frequent in the education literature, some introductory sequential analyses have been conducted that have experimented with this type of analysis. Examples include implementing a sequential method to describe various levels of teacher proficiency (Hawkins & Sharpe, 1992), to provide feedback+goal setting in undergraduate practice-teaching settings (Sharpe, Hawkins, & Ray, 1995), and to increase the self-monitoring skills of practicing teachers (Sharpe, Spies, Newman, & Spickelmier-Vallin, 1996). In addition, some experimentation has been conducted in the area of alternative variable construction in an attempt to make the interactive transactions between teachers and students in challenging instructional and organizational situations more explicit (Sharpe & Lounsbery, 1998; Sharpe, Lounsbery, & Bahls, 1997).

With the software tools contained with these materials, we have tried to provide applications that will facilitate a wide range of behavior-behavior and behavior-environment analyses and accommodate a host of different types of category system constructions depending on the variables a particular professional is interested in describing and analyzing and depending on the type of information (or measurement) that professional wishes to implement. In this regard, this chapter provides a general framework for category system construction, which may be helpful to users developing their own particular direct observation systems for their own particular needs, and provides some specific category systems that have been used with success for various research and development and related evaluation purposes in various education and social science situations.

Many nontraditional coding systems are also available in the literature, which may provide additional insight as well and include such dimensions as power and involvement (Penman, 1980), control and support (Pett, Vaughan-Cole, Egger, & Dorsey, 1988), and emotion dynamics (Ekman & Friesen, 1978). For a detailed discussion of category system construction and the conceptual considerations one may wish to consider, we also recommend a thorough reading of Bakeman and Gottman (1986) as they provide what we feel to be an excellent treatment of these issues.

Clearly, the construction of a category system that best fits the description and analysis task one may wish to undertake is the scientific step requiring the most thought, deliberation, and reflection. Because using behavioral codes to collect and analyze behavior-behavior and behavior-environment data often reduces the richness of the interaction to a finite set of categories, the choice of a coding system that is sensitive to the form and character of the interactions among setting participants is critical. Because of this potential limitation with regard to using behavioral codes, we felt it important to also include a qualitative narrative data collection feature within the data collection tools provided (refer to the technical guide materials).

Some Cautions

Prior to a detailed explanation of the important considerations regarding category system construction for a particular use, some general cautions should first be highlighted. We recommend that the reader remain sensitive to these cautions not only throughout the category system construction process but also throughout all data collection and analysis efforts.

A first caution relates to the primary intent of "doing" science. If scientific practice is focused on providing an inclusive and accurate description of the world around us, and this practice operates under the rubric of field theory, then all the software tools that we provide succeed in accomplishing is quantitative, qualitative, and sequential *description*. The question that is answered by the quantitative data provided is how often and how long certain behaviors and events tend to be present or absent in certain situations. The questions that are answered by the qualitative narratives relate to contextual and situational features that surround the presence of particular behaviors and events. The question that is answered by the sequential data is whether behavior B follows another behavior A more or less often than would be expected by chance. If the frequency of behavior B following behavior A is sufficiently different from chance occurrence (as determined by the *Z* score–based mathematical models), then a probabilistic dependence is present between behavior A and behavior B. Whether focusing on the quantitative, qualitative, or sequential characteristics of particular behaviors, it would be unwarrantedly optimistic for a user of our tools to state unequivocally that behavior A *caused* the occurrence of behavior B, for causation is a separate epistemological question and one that we feel important to separate from scientific description and delineate from a determination of probable dependence.

Assumptions

Regarding the collection and analysis of traditional quantitative data (e.g., discrete characteristics of behaviors and events such as frequency, rate, or percentage of experimental time), some assumptions are important to note. First, just because a category system is used to guide the quantitative collection of behavior-environment data, the user should not assume that the reliability or accuracy of that data is ensured. It is in this respect that a rigorous set of rules and procedures should be implemented to increase the probability of reliable and accurate data records. A reasonably complete treatment of some recommended procedures are provided in Chapter 4 (refer to relevant pdf file on your CD ROM disk) and should be given a thorough read by those less familiar with recommended reliability and interobserver agreement methods.

Second, we agree with van der Mars (1989) that while it is assumed in most methodological literatures that a category system approach concentrates on observable behaviors and events, such a method must not necessarily be assumed to be capable of only measuring those events that can be detected visually or audibly. Clearly, we agree with Friman, Wilson, and Hayes (1998) in our feeling that it is a misconception to automatically rule out this type of method when interested in phenomena such as emotions, feelings, attitudes, and perceptions. Pioneers of behavior analysis (see Skinner, 1989), for example, have made it clear that most feelings and attitudes (labeled private events) may be connected with a readily observable manifestation of those emotions. van der Mars provides an education illustration of this as follows:

> If we assume that attitudes and feelings are somehow going to be reflected in observable behaviors, such behaviors can be categorized and defined. For example, attitude of an athlete toward practice might be reflected in his or her on-time behavior, the amount of time he or she spends taking extra practice, and so on. Although these and other behaviors may not say much about that person's general attitude, they function as indicators of his or her attitude toward practice. (p. 8)

Related to the sequential nature of the data collection and analysis tools we provide, Wampold (1992) lists some important assumptions. A first assumption is that the probability that event A is emitted is independent of its position within the larger behavior-event sequence of interest and, according to the null hypothesis of randomness, is independent of the other behaviors in the event sequence to be studied. This assumption is referred to in much of the sequential literature as "stationarity."

Stationarity implies that the patterns of interaction will not change over the course of the observation. However, this may not be true in all sequential interactions, particularly those among various participants in education settings. For example, a teacher may change his or her interactional pattern within a particular class period according to the lesson context and the different skills being taught. When analyzing a sequential pattern in which stationarity is not uniformly present, it is recommended that the behavior-behavior or behavior-environment sequence of interest be broken up into stationary segments and analyzed as independent data records. For example, introductory and closing behavioral exchanges in an instructional lesson may be segmented from a sequential analysis of the interactions present during the body of that lesson. In addition, as the body of the lesson changes contexts (e.g., from seat work to cooperative activities), the data record should be segmented accordingly.

A second assumption by Wampold (1992), and related to his first assumption that behaviors are independent of other behaviors in the sequence, requires that behaviors under observation not be constrained structurally in any way by the method of data collection. One example of a common constraint in the more traditional behavior analysis literature in education includes sequences that are collected by a data mechanism that does not allow a behavior or event to be followed by itself. This type of sequence is collected when a mutually exclusive coding system is used that only records a behavior when a change of behavior occurs. For example, oftentimes a teacher's instructional behavior will be recorded until that teacher changes to another behavior such as management or observation. A teacher's instruction is therefore not allowed to follow a teacher's instruction, and misleading conclusions may result (see Wampold, 1986, for a more detailed discussion of this issue).

Limitations

A potentially limiting methodological characteristic brought up most often when discussing the relative merits of a category system approach to the observation of behaviors and events is its inability to provide prescriptive information. Similar to qualitative methods, once a behaviorally based descriptive data record has been compiled and analyzed, what a particular researcher or clinical supervisor does with that information is a function of that professional's interpretation. In other words, it is important to note that the data record in and of itself does not provide a prescriptive interpretation of the relative effects of behavior or environmental events. The data record only provides descriptive, and hopefully accurate, information on which analytic and prescriptive judgments will be made. Clearly, this method for collecting and analyzing data does not alleviate the need for a trained and experienced professional to interpret the results according to recommended procedures. It is in this regard that it is important to adhere to a strict set of procedures when interpreting behavior-environment data to avoid superimposing the beliefs and biases of the observer on the situation under analysis. For a detailed discussion of recommended behavior-environment data interpretation, refer to the more complete behavioral assessment and data evaluation treatments provided by Kazdin (1982) and Sulzer-Azaroff and Mayer (1991).

A second potential limitation of this type of work lies in the mischaracterization that is oftentimes promoted when this method is used for evaluation purposes. A teacher-training example perhaps illustrates this point most effectively. Many professional teacher educators have argued against a category system approach to the quantification of various teacher and student behavioral practices due to its perceived technocratic character. In other words, a technocratic argument promotes that a behavioral observation is inappropriate due to the fact that it provides rigidly defined behavioral competencies that are either displayed to a proper amount or not displayed to a proper amount in a general way. The argument then follows that teachers-in-training are therefore inappropriately held to a rule-governed approach to either demonstrating particular behaviors at a particular level of performance (e.g., minimum frequencies or percentages of class time) or failing their evaluation experience.

Clearly, however, the data records generated by this type of approach are contextually bound. We caution against using the type of quantitative data records this method generates to stipulate minimum levels of teacher behaviors or student practices related to instructional effectiveness, for example. What is recommended when using this type of data for evaluation purposes is that the data record should be considered with respect to the situation in which it was observed and the background and experiences of the participants who were under observation. For example, it would be inappropriate to expect the same levels of behavior from a novice teacher who rotates to different school settings on a daily basis as those of an experienced teacher who practices in the same setting on a regular basis. It would also be inappropriate to expect the same behavioral levels from the same teacher as he or she taught across different subject matters with different classroom arrangements, resource needs, and activity requirements. Hence, and related to our first assumption, the degree to which the data generated according to this method are helpful to particular clients is reliant in great degree on the trained experience of the professional who interprets that data for various evaluation or research-driven ends.

Given our enthusiastic arguments in favor of a sequential approach to the collection and analysis of observed behaviors and events, some limitations specific to the sequential dynamics of the software tools provided should also be highlighted. First, the construction of an amenable observation system to use for the coding of behavior and the actual data collection process consumes time and technological resources. Only a few computer-based software systems are currently available to help in this challenge, and of those most are not available commercially. Therefore, resource availability in relationship to the hardware needed to support such an approach must be carefully considered before implementing this type of system. While we argue that computer hardware and the operating systems that are typically included with most hardware are becoming increasingly capable and increasingly less expensive, the expense issue remains important to consider.

Related to the expense limitation is that when conducting research into the complex behavioral character of certain settings of interest, collection and analysis from videotape permanent records are methodologically desirable and in some cases necessary. Complex overlapping coding schemes require multiple passes through a permanent record to capture completely the data of interest. In addition, when training a large staff for reliability and consistency of data collection efforts, videotape records and multiple computers are required to implement such an undertaking (refer to Chapter 4 pdf file on CD ROM for a more detailed procedural discussion of these issues). Despite potential equipment limitations, once coding system construction and reliability issues are addressed, a sophisticated category system approach to the quantifying of observable events is easily undertaken by even the relatively inexperienced researcher due to the current sophistication of data collection and analysis programs such as that contained in these materials.

A second limitation related to the sequential nature of our data collection and analysis recommendations is that this type of analysis should be limited to behavior-behavior and behavior-environment interactions that occur with relatively high frequency and occur with relative immediacy. Important events that occur infrequently or very irregularly are best left to another method.

A related, and equally impacting, issue is that of determining the unit of analysis (e.g., the entire class, small groups of students interacting with one another, one problematic student, etc.). If the research or evaluation task using the tools we recommend is not contextualized by a particular unit of analysis (refer to Silverman & Solmon, 1998 for a detailed treatment of this issue), then results from such an activity may not yield meaningful, in the best case, or may yield confounding, in the worst case, results.

Implications

Despite the cautions we have presented, an observational approach to the discrete and sequential counting of behaviors and events provides some appealing benefits. First, a tool is provided for discovering, documenting, and quantifying how behaviors tend to be related to one another in situations largely composed of interaction among many individuals. Education and social science settings, for example, in which the predominant component is the interaction among multiple individuals in a particular context seems an ideal match with such an analysis. A more capable descriptive tool that these materials provide should also prove productive in overcoming the often voiced constraint of undue focus on a limited number of behaviors or events in isolation. Providing a qualitative data collection component should also help overcome the sometimes wrongly assumed context-free nature of those behaviors and events.

The capabilities that this approach provides also seem well suited for uncovering many of the more subtle behavior interactions unique to particular participants. Characteristics such as automaticity, contingency management, and response time, all thought to possibly be components unique to very effective interactions, may now be amenable to documentation and quantification. Key interactive characteristics such as rapidly changing rates and latencies of participant responding also may be uncovered within and across specific settings. Once documented, these more subtle and complex behavior-behavior and behavior-environment configurations may be amenable to successful training and transfer, thereby improving the effectiveness of various professional practices in education and social science settings.

Given the capability of the software tools that we provide, the level of complexity of quantitative and qualitative description and analysis is only limited by investigative interest and the inventiveness of the category system developed to collect corresponding data. Once decisions such as determining appropriate units of analysis, overcoming the assumptions of behavioral independence, and gathering the required resources are met, the appeals of this type of data gathering and data analysis are many. What is now necessary is providing some general guidelines along with some specific examples and illustrations to get started on actual category system construction for a particular application.

Construction Recommendations

When beginning the task of putting together a set of terms and definitions to form a category system to be used for direct observational purposes, there are many things to consider. The construction process is particularly important, again, as it is the terms and definitions of a particular category system that will act to guide all remaining data collection and analysis efforts.

Defining Purposes

A logical first step to actual category system construction is to first ask those who will be using the category system for various tasks what the general purpose or goal of the projects in which the category system will be used entails. If the general goal is evaluation, then discussion needs to take place regarding what the most important features of the instructional or training experience are. This is an important step, as the category system to be developed is ultimately designed to document whether particular clientele behaviors are indicative of successful instructional or training effects. In other words, the purpose of recording behaviors and events in an evaluation effort is to categorize the important features of the education or training experience into distinct categories for data recording. This is an important step in that the terms and definitions contained within the category system will be indicative of what various clientele will be held accountable for as a function of exposure to a particular education or training experience. Careful thought concerning terms and definitions in relationship to larger education or training issues ensures a clear connection between the larger emphases of the training program and the behavior and environment events contained within the related category system—events that clientele will be held accountable to demonstrating in certain ways as a function of being successfully trained.

The same idea holds true for a research and development enterprise that uses a category system for the collection and analysis of behaviors and events of interest. In a research and development context, however, the purpose question is twofold. First, a set of terms and definitions should be provided that are capable of describing as completely and accurately as possible the educational treatment or experimental intervention that is being tested for its relative effects. Whether the treatment is systematically manipulated (e.g., introduced or withdrawn at specific intervals or points in time in a particular setting) or exists as a natural part of an ongoing setting of interest, at issue is the importance of being able to completely and accurately describe the presence or absence of the treatment and/or its various component parts. The descriptive recording of the characteristics of a particular treatment is an important and, until recently, an often underemphasized dimension to the observational research process. In this regard, we want to gain a clear understanding of whether the treatment was implemented according to the recommended procedures or practices set up as part of the experiment. If the treatment is not, and certain treatment variations are not recorded, then what is often termed the *fidelity* of treatment intervention, or treatment integrity, is severely compromised. If treatment fidelity is compromised, then analysis results may also suffer compromise in terms of inaccurate or incorrect experimental conclusions. Witt, Noell, Lafleur, and Mortenson (1997) provide a contemporary data-based illustration of how one might measure and analyze treatment integrity in an education setting, and detailed discussions of this issue are available in the applied behavior analysis literature for those interested in additional information on this subject (see Gresham, Gansle, & Noell, 1993; Peterson, Homer, & Wonderlich, 1982).

The second purpose question related to a research and development activity relates to the measures chosen to determine whether a particular treatment or intervention was effective. In this regard, it is important to construct a set of terms and definitions that will most accurately and inclusively represent what you as a researcher hold most important regarding the desirable effects of a chosen intervention. This is the case whether you are looking for evidence of the presence of or increase in certain behaviors or events as a result of a particular intervention or whether you are looking for evidence of diminishing or absent behavior or event occurrences. Care in the construction of measurement-oriented terms and definitions is also important whether the research question involves determination of the relative effects of a manipulated treatment or whether the question involves the effects certain behaviors and events may have on certain others in settings in which both “treatment” and “measure” are naturally occurring.

Determining System Characteristics

At the same time that a general set of terms and definitions are being outlined for the purpose of constructing a definitive category system, some structural questions must also be answered in relationship to how a particular category system code is organized to capture those behaviors and events of interest. Code organization is fundamentally related to the type and scope of information a researcher or evaluator wishes to gather in relationship to other behaviors or events of lesser or negligible interest.

When determining system structure, three basic decisions must be made. The first relates to whether the terms and definitions to be used for collecting data will be mutually exclusive of one another or whether certain behaviors and/or events will be recorded at the same point in time (i.e., potentially occur in concert or overlap in occurrence). For behavior and event categories to be mutually exclusive, it simply means that if one of the behaviors or events in the category system is being recorded, then all (or a specified subset) of the other behaviors and events contained within the category system being used must not be recorded as occurring at the same time (refer to the data collection sections of the technical guide for a detailed discussion of how the software tool may predefine a set of rules to guide the mutual exclusion of certain events if certain other events are being recorded). Making such a distinction is important depending on the type of data-based information that is desired.

Collecting data in a movement education classroom provides an illustration of this point. If, for example, you are interested only in how certain types of instruction affect the relative success of student practices when the teacher is proximate to a student, you might opt for a mutually exclusive code. In this context, you might logically separate verbal instruction from skill modeling and separate these two instructional behaviors from the instructional behavior of physical guidance. Though modeling may also be defined to contain elements of verbal instruction, and physical guidance may be defined to contain elements of both verbal instruction and/or modeling, in this example, data collection interest is on the relative proportion of each instructional behavior as a separate entity and should therefore be recorded separately. If, on the other hand, you are interested in describing the various ways in which certain types of instruction (e.g., verbal, modeling, and physical guidance) may overlap with one another to form more effective or meaningful combinations of instruction for particular students, then these behaviors should be defined to allow the coding of more than one instructional behavior when more than one of these behaviors occurs in concert or in overlapping fashion.

Another issue related to mutual exclusivity is more of a pragmatic matter and concerns whether a category system is being used to record data directly in a live setting as the behaviors and events naturally occur or whether data is being recorded from videotape records of the setting of interest. Clearly, if recording live, a mutually exclusive format is easier to implement. Using a mutually exclusive structure will heighten the probability of an accurate data record, for only so many keys on a computer keyboard may be humanly pressed at the same time and only so many behaviors and events (keys) can be recorded with an acceptable level of accuracy as multiple events occur in concert or overlap with one another. In the live recording case, the complete and inclusive nature of the data record is oftentimes sacrificed to a certain extent for the sake of feasibility and accuracy concerns. On the other hand, if recording data from videotape records, a much more complex and overlapping event record is feasible, one that may provide a much more accurate and inclusive representation of the multiplistically overlapping nature of an observed setting's behavior-behavior and behavior-environment interactions.

A second system decision relates to whether the category system needs to be capable of an all-inclusive (or "exhaustive" as termed by Bakeman & Gottman, 1986, p. 33) description of any potential behavior or event occurrence within a setting of interest or whether particular behaviors are the primary focus with all others of lesser importance considered so to the point of not being specifically included in the category system. For the purposes of this discussion, a category system is termed all-inclusive if any behavior or event that may conceivably occur in the setting being observed can be recorded as one of the specified categories within that system.

Examples from education again serve well to illustrate this decision. Suppose, for example, a clinician in a resource setting for the severely and profoundly autistic and self-injurious clientele is interested in the effects of a particular set of instructional behaviors on the incidence of self-injurious episodes. If this is the focus of a data record, then a category system that includes (a) particular instructional behaviors, (b) the primary measure of self-injurious incidents, and another broad category of (c) other might be a most appropriate system configuration. In this way, all other behaviors and events in the setting may be lumped together in a general or anonymous way to retain the time-based integrity of the data collection episode, but specific behaviors and events of interest are focused on for recording purposes using a less than completely inclusive system. Suppose, on the other hand, the same clinician is interested in describing the many possible interaction effects of all behaviors and environmental events that are operating within the same resource setting for the purpose of determining which behavior-behavior and behavior-environment interactions might be relatively more or less therapeutic in the reduction of self-injurious behavior. If this is the case, then a more all-inclusive category system should be developed that provides explicit information regarding *all* of the behaviors and events operating in time-based proximity to self-injurious events within that resource setting.

A third system-level decision relates to the size and complexity of a category system regarding the number of actual terms, or codes, contained within the system to be used. This decision is also integrally related to whether data collection will take place live in the setting in which behaviors and events of interest are occurring or will take place from videotape records of that setting. While a structured set of rules and procedures does not explicitly exist as they do for various data analysis or reliability guidelines referenced elsewhere in these materials, we recommend some general rules of thumb. First, we agree with Bakeman and Gottman (1986) that it is always prudent to err on the side of simplicity rather than undue complexity when constructing a category system. An inherent danger to being capable of describing a host of behaviors and events as they occur is that meaningful information may be lost in a proverbial sea of data if the category system is unduly complex in terms of the number of categories included for recording purposes. Though this is clearly an area in which further study is necessary, our experiences have convinced us that once we go beyond a certain number of individual categories in a particular system, the marginal utility of the amount of information gained becomes increasingly less than the increasing challenge of ensuring the accuracy and reliability of the data record. In other words, a danger of infinite regress into potentially less and less meaningful discriminations among very similar behaviors and events clearly exists as a category system becomes more complex. In our experience, we have found that for live recording, after going beyond a limit of 14 to 16 categories within a particular system, the quality of information ceases to meaningfully increase, and the feasibility of recording information accurately and reliability drops off sharply.

In contrast, we have been enamored with the ability to record from videotape records an increasing number of behaviors and events with finer and finer discriminations among behavior and event classes when using a particular category system. Through various methods (e.g., collecting and merging separate data records containing logically grouped behaviors and events—refer to the data merging sections in the data analysis discussions contained within the technical guide), it is even possible to ensure the accuracy and reliability of the data records gathered from videotape when using a very complex category system. The computer-based tools we provide allow for 36 categories and the subgrouping of each of those 36 primary categories into 99 subcategories by using numerical notations within each primary category. Even though we are capable of such data recording complexity when using the computer-based tools provided with these materials, we have found that the same law of diminishing returns is inherent to recording data from videotape as from recording live. We have typically found that after expanding a particular category system beyond 25 to 30 events, the amount of meaningful information gained from doing so is quite minimal when recording from videotape. For example, including a group of 4 or 5 questioning behaviors to a category system designed to study effective instructional practice in postsecondary settings not only may prove very insightful to which types of questions are most effective in facilitating productive student interactions but also may prove equally helpful to determining the most effective sequencing of questioning behavior for optimal instructional practice. If we were to split these 4 or 5 questioning categories further into say 15 or 20 subcategories by the inclusion of various verbal and nonverbal cues and various other contextually based dimensions for each primary questioning category, we may not gain much additional information in the typical postsecondary setting under observation.

It is also important to provide a final system decision note regarding our software tool's capability to collect and analyze behaviors and events in multiple ways. The software provided includes, for example, multiple analysis capabilities with respect to (a) the individual or discrete character of behaviors and events, (b) the sequential or time-based nature of event occurrences in situational context, and (c) the qualitative character of the behaviors and environmental events under observation. Beyond the technical decisions of how to construct a category system and actually record behaviors and events (i.e., the press-and-hold, toggle, numerical notation, comment features, and so on described in the data collection sections of the technical guide) in the most economical way, research and evaluation decisions must also be made regarding the most informative measure in relationship to the observational purposes decided on for a particular application. In this regard, we have provided the capability within our software tools (ensured by the conceptual nature of how data records are constructed as behaviors and events are recorded and ensured by the relatively inclusive nature of the descriptive measures contained in the various analysis applications) to analyze each behavior and event recorded across all of the measures contained within the software tool. What remains, however, for the user is the challenge of reporting the most salient measurement data to a particular research audience to get their point across and reporting the most helpful measurement data to various clientele when using these tools for evaluation and feedback purposes. As very detailed discussions of this issue are available elsewhere, we recommend reading Bakeman and Gottman (1986); Barlow and Hersen (1984); Cooper et al. (1987); Johnston and Pennypacker (1980); Kazdin (1982); Kerlinger (1986); LeCompte and Preissle (1993); Miles and Huberman (1984); and innumerable other behavior analytic or qualitative methodology texts for a thorough treatment of measurement choice issues.

On the Construction of Definitions

A last important consideration when constructing category systems for observational purposes lies in just how to articulate the definitions of the terms included in the system. We agree with Kazdin (1982) that behaviors and events included for observational purposes need to be defined very explicitly so that they can be "observed, measured, and agreed on by those who assess performance and implement treatment" (p. 23). Successful description, and accordingly successful analysis, begins, according to Hawkins and Dobes (1977), with objective, clear, and complete articulation of behaviors and events of interest. In their view, *objectivity* refers to limiting definitions to the observable characteristics of specified behaviors and events. If truly objective, then definitions should avoid all references to behavioral intention or to internal states or mentalistic perspectives on those behaviors (Barlow & Hersen, 1984). *Clarity*, on the other hand, refers to the ability of different observers who use the category system definitions to read, interpret, and articulate instances and noninstances of particular behaviors and events with a lack of ambiguity or disagreement. Finally, the characteristic of *complete* refers to ensuring the articulation of boundary conditions that delineate behaviors and events across categories in the observational system. In other words, a set of category system definitions is said to be complete (and clear) if it is plain and apparent to multiple observers how certain behaviors and events should be included or excluded from a particular category.

The characteristic of complete with regard to defining behaviors and events in a category system is perhaps most important. To ensure the complete nature of behavior and event definitions, a set of rules needs to be generated that makes explicit how particular behaviors and events should or should not be recorded. According to Hawkins (1982), to ensure that a definition is complete, it must include all of the following:

1. A descriptive name
2. A general description of the behavior or event much like that found in a dictionary
3. A discussion of the critical components of the behavior or event that may be used for rulings on inclusion or exclusion
4. Typical examples of the behavior or event that help in making questionable or borderline judgments on difficult calls when recording

We have found that one helpful strategy to ensure an objective, clear, and complete set of category system definitions is to watch situations in which behaviors and events of interest occur often and to first form an inductive and general set of descriptions of each targeted behavior and event. Next, we recommend collaboration with others in coming up with examples that are helpful in making borderline judgments on potentially ambiguous occurrences of each behavior or event in question. If this procedure is followed, it is less challenging to finalize a set of definitions that meet the criteria of Hawkins (1982) as listed.

Despite what may seem to some like a rather overwhelming list of methodological cautions, and a relatedly daunting list of category system construction decisions and definitional recommendations that we have provided, we remain confident that a direct observational approach to the more inclusive quantifying of behaviors and events holds great promise and appeal. Though we have provided a representative list of references of various education and social science direct observation applications, we felt that it would be helpful at this point to provide a general terms and definitions framework and some specific category system examples that emanate from that more general framework to help those using these materials construct their own category systems for their own particular needs.

Examples and Illustrations

The following sections provide the new user with some select observation instrument examples that have been used with some success in various education and social science situations. Some of these examples are quite detailed and offer illustrations of terms, definitions, and referenced lists of evaluation strategies based on the use of that category system as a feedback instrument. Other examples are limited to a summary of category terms as more detailed illustrations are available in our list of references, and the definitions to these terms are fairly straightforward. All examples are provided in the hope of stimulating thinking about the types of terms and definitions that may be best suited to a particular research or evaluation task, and for this we thank our various contributors. As stipulated in the technical guide, the first two education-based configuration files (i.e., PETEACH and ELED1) or category system examples contained here are also contained in file form to be accessed and used when becoming familiar with the data collection software application.

A General Framework

The following framework is provided to stimulate thinking across the many different possible behavior and environment classes that might be important to include for various research or evaluation purposes. Some classes are time and behavior based, and some are not but remain important to a complete and inclusive description of an experimental or evaluation setting. By providing this framework as a general guide to stimulate constructive thinking, we hope that specific category systems may be more readily designed and more productively implemented.

Table 1 General Framework

Classes	Description	Examples
Environment		
Setting	Service delivery setting or research setting	Classroom, resource room, natural habitat, or client environment, and so on
Content	Subject matter to be taught, treatment or intervention to be introduced, skills to be learned	Math or reading skills, positive social treatments, behavior reduction strategies, and so on
Content stage	Temporal status of the observational setting	Beginning, middle, or end; lesson preview, body, or review
Materials	Physical resources	Educational equipment and materials, socially stimulating toys, client workbooks or folders, and so on
Participant grouping	Physical arrangements	Large or small group, number and kind of participants, and so on
Method of interaction	Stimulus method to occasion responding	Instructional or interactive styles such as command, individual, or reciprocal
Transitions	Events signaling situational changes	Changes in participant grouping, in educational or treatment context, in physical arrangements, and so on
Teacher, therapist, or leader behavior		
Behavior	Relative to student, client, or other setting participants	Observation, verbal and nonverbal instructions and directions, questioning, responding, organizational and interpersonal interactions, compliance, resistance, and so on
Physical movements	Relative to position in setting	Sitting, standing, moving, reaching, waving, eating, sleeping, walking, bar pressing, physically guiding another, and so on
Focus	How behavior and movement is directed	To an individual or defined subgroup of participants
Setting position	Relative proximity to other defined setting participants	Central or peripheral setting locations, proximate or distant to defined setting participants, presence or absence in setting
Student, client, or participant behavior		
Behavior	Active stimuli and responses	Activity engagement (successful and unsuccessful; desired and undesirable), supporting and instructional behavior, listening, waiting, off-task, disruption, responding, interpersonal, and so on
Physical movements	Relative to position in setting	Sitting, standing, moving, reaching, waving, eating, bar pressing, physically guiding another, and so on
Focus	How behavior and movement is directed	To an individual or defined subgroup of participants
Setting position	Relative proximity to other defined setting participants	Central or peripheral setting locations, proximate or distant to defined setting participants, presence or absence in setting
Historical		
Contextual	Past and present setting factors affecting behavior performance	Physical arrangements, participant arrangements, setting materials, methods of interaction, and so on
Experiential	Participant background	Education, years and types of experience, cultural background, age, socioeconomic background, interactive history

NOTE: Adapted from Hawkins, Sharpe, and Ray (1994).

Teacher Education

This section provides two category systems defined according to more traditional and more frequently seen systems in education. Each is typified by mutually exclusive and all-inclusive system decisions, and each was designed for evaluation purposes when collecting data live in actual practice-teaching settings within a teacher education program. The first category system is designed for use in physical education and youth sport settings in which movement and active participation are desirable student behaviors and a variety of instructional behaviors are used by both teacher and student. A list of definitions is also provided with this example to give the reader some illustrations of definitions that meet the objective, clear, and complete criteria we have discussed. Of additional appeal, rich lists of feedback strategies are available in the literature in relationship to this type of category system when used as a feedback and goal-setting tool in the preparation of physical education teachers and youth sport coaches. Referring to these references may help those using our software tools for evaluation purposes by illustrating the ways in which professionals might construct a complimentary set of educational information for various clientele based on the behaviors and events contained within the category system constructed for evaluation purposes. For those interested in a detailed discussion of the feedback strategies used in concert with this type of category system in relation to teacher training procedures, refer to Hawkins, Wiegand, and Landin (1985) and Landin, Hawkins, and Wiegand (1986). This category system is contained in the software program files as a data collection configuration system labeled PETEACH. For additional discussion of this category system and how it was derived, refer to Hawkins, Sharpe, and Ray (1994).

For the second category system illustration, we provide the behavior and event terms only. This example was designed for use in evaluating elementary education teacher trainees in public school classroom settings. Only the terms are provided to illustrate in summary form how an all-inclusive category system might be constructed to serve evaluation needs when providing feedback information to teachers practicing in elementary education classroom settings. This category system is also contained within the software program files as a data collection configuration system labeled ELED1. This category system is based in part on Stallings et al. (1987), which should be referred to for a more detailed discussion of some of the definitions contained within this category system and how this system has been put to use.

For each category system illustration, the actual numbers and letters used to collect data with the computer keyboard are specified next to each behavior or event term to give a sense of how each category system is organized for actual data collection.

Table 2 Physical Education Teacher Education (PETEACH) Category System Summary

Teacher Behaviors	Student Behaviors
1. General observation	M. Motor appropriate
2. Specific observation	M-F3-1-RETURN. Motor engaged
3. Encouragement	S. Supportive
4. Positive feedback	C. Cognitive
5. Negative feedback	N. On-Task
6. Management	F. Off-Task
7. Verbal instruction	I. Instruction of peers
8. Modeling	W. Waiting
9. Physical guidance	
0. Interpersonal	
A. Off-Task	

Contextual Elements (toggled keys in PETEACH)

T. Transition
P. Preview
B. Lesson body
R. Review

NOTE: Adapted from selected category systems in Darst, Zakrajsek, and Mancini (1989).

PETEACH Definitions

Teacher

1. General observation: The teacher is watching student groups engaged in any category of student behavior. This category includes passive supervision, and there is no relationship of the observation to an instructional focus. The teacher must also *not* be engaged in any other category of teacher behavior to record general observation.

2. Specific observation: The teacher is watching *one* student engaged in a subject matter task for the purpose of providing feedback related to performance. The teacher position must be proximal to the student position so that observation is clearly focused on a specific student who is performing. Specific observation could also be recorded when the teacher is watching pairs or small groups of students when the instructional focus is clearly on a group task (e.g., observation of five players executing a fast break during instruction on the fast break in basketball).

3. Encouragement: The teacher makes a verbal statement *prior* to a student skill or organizational attempt, which is clearly to enhance the student's perception of his or her ability to accomplish the subsequent task. The teacher is not telling the student what to do (e.g., an instructional prompt - behavior 7) but is clearly trying to build confidence (e.g., "you can do it," or "if you did it last time, you can surely do it this way," etc.). This category may also be recorded when encouraging behaviors are conveyed to the class population as a whole or to small groups of students.

4. Positive feedback: The teacher makes a positive verbal statement or gesture following an individual's or group of students' skill or organizational behaviors, which is clearly designed to increase or maintain such responses in the future. The statement or gesture must follow soon enough after the behavior that the student clearly associates it with the behavior commented on. Feedback statements may easily be delineated from encouraging statements, for encouragement occurs prior to the student behavior in question and feedback occurs after.

5. Negative feedback: The teacher makes a negative or critical verbal statement or gesture following an individual's or group of students' inappropriate skill or organizational behaviors, which is clearly designed to decrease or eliminate such responses in the future. The statement or gesture must follow soon enough after the behavior that the student clearly associates it with the behavior commented on.

6. Management: The teacher is engaged in carrying out a non-subject-matter organizational task (e.g., setting up equipment, taking roll, collecting papers, explaining station rotations, directing students to move to another instructional activity or station, etc.). This category may be conducted in a verbal or nonverbal gesturing manner.

7. Verbal instruction: The teacher is verbally describing to the students how to do a skill or is using a verbal prompt to direct a student or group engaged in attempting a skill or activity. The student task must be a subject matter activity to record verbal instruction.

8. Modeling: The teacher demonstrates to students how to do a subject matter task or participates with students in a subject matter task or activity. If the teacher uses a student to demonstrate a subject matter task, this category should also be recorded for the duration of the student demonstration episode.

9. Physical guidance: The teacher physically guides an individual or group of students through a subject matter task or activity. Actual physical contact must be made and maintained with the student in question for this category to be recorded (e.g., holding a student's arm on a balance beam in gymnastics, guiding an athlete's arms through a proper swimming technique, etc.).

0. Interpersonal: The teacher talks to an individual or group of students about non-subject-matter and nonmanagerial tasks in a manner that is clearly designed to foster a positive interpersonal relationship between teacher and student. Commenting on a student's clothing or talking about what one student did over the weekend are examples of interpersonal.

A. Off-Task: The teacher is clearly not paying attention to the instructional and/or organizational responsibilities regarding the class at hand. Making notes on what to do during football practice during the course of a physical education class, flirting with the passing office staff, or daydreaming against the gymnasium wall are clear examples of off-task behavior.

Student

M. Motor appropriate: The student is engaged in a subject-matter motor activity in a successful manner. Success is defined as meeting the lesson objectives of the teacher or coach. Examples include dribbling around cones with a basketball without letting the ball get away, correct skill performance as defined by a coach in a practice setting, and so on.

M-(Number Notation - 1). Motor inappropriate: The student is engaged in a subject-matter-oriented motor activity, but the task is either too difficult for the individual's capabilities or the task is so easy that student practice is performed poorly or incorrectly, clearly not meeting lesson goals. Examples include dribbling around cones with a basketball and the ball gets away from the student, shooting free throws in a basketball practice session but missing the basket, doing improper hand turns during a swimming practice set, and so on.

S. Supportive: The student is engaged in assisting others to perform a subject matter motor activity (e.g., spotting in gymnastics, feeding balls to a hitter in a tennis lesson, throwing a volleyball to a partner who is practicing set-up passing, clapping a rhythm for a group of students practicing a dance movement pattern, etc.).

C. Cognitive: The student is attentively listening to the teacher or a visual aide explain an organizational or subject matter task (e.g., verbal description of a game, watching a modeling episode, viewing a filmstrip, participating in a discussion, etc.).

N. On-Task: The student is appropriately engaged in carrying out an assigned non-subject-matter task that is designed to *prepare* for a learning and/or skill attempt (e.g., moving into squads, moving from the gymnasium to the playing field, reading prescription sheets at a drill station, etc.). This category may be equated with any student managerial tasks undertaken to attain a state of learning readiness.

F. Off-Task: The student is either not engaged in an activity in which it is clear he or she should be engaged in or is engaging in an activity other than the one clearly advocated by the teacher (e.g., behavior disruptions, misusing equipment, fighting, etc.).

I. Instruction of peers: The student is clearly teaching either an individual or group of his or her peers regarding the subject matter activity at hand. This category includes student performance of any of the three teacher instructional behaviors (e.g., verbal instruction - 7, modeling - 8, or physical guidance - 9).

W. Waiting: The student has completed a task and is awaiting the next instructions or opportunity to respond. Waiting in line for a turn, waiting for the next teacher direction, waiting to get into a game from the sideline, waiting for the next activity to begin, and so on, are all examples of this category.

Context

T. Transition: The setting structure of the education or sport setting changes from one form to another. This event should be recorded for the entire duration of the transition event to provide information related to length of transition and total evaluation time spent in transitions. Examples include student movement from skill station to skill station during a lesson, movement from skill practice to a lead-up game situation at the end of a practice session, or movement from warm-up activities to a skill practice activity.

P. Preview: This category indicates that the lesson is in its introductory stage. This time period encompasses the initial portion of the class or practice experience. Elements that allow the class to attain a state of lesson body readiness are involved and typically include taking roll, reviewing for students what has gone on in the previous lesson, initial teacher-directed organizational statements, verbal encapsulations of what is to be encountered in the upcoming lesson, warm-up exercises, and so on. Ideally, this contextual phase should only take up the first 3 to 10 minutes of class time.

B. Lesson body: This category indicates that the lesson is in its main instructional phase. This time frame encompasses the middle portion of the class or practice period in which subject matter content instruction is central to the classroom focus. Station and/or drill work, lead-up games, teacher- or pupil-based instructional time, and so on, typically make up this contextual event.

R. Review: This category indicates that the lesson is in its final closure stage. This time period encompasses the final minutes of the class or practice experience. Elements typically include teacher-directed review of the major points of the lesson just encountered, tie-in statements with previous lessons or related subject matters, verbal encapsulations of what is to be encountered in the next class or practice meeting, final comments to individual students or groups, and so on. Ideally, this contextual phase should only take up the final three to five minutes of class time.

Table 3 Elementary Education Category System (ELED1) Summary Example

Teacher Behaviors	**Student Behaviors**
1. General observation	S. Content Work
2. Specific observation	Seat activity
3. Encouragement	Game activity
4. Verbal instruction	Board activity
Information	L. Listening (passive)
Concepts	Q. Questioning
Questioning	R. Responding
6. General praise	H. Helping
7. Content-specific feedback	Peers
8. Management	Teachers
9. Reprimand	O. Organizing
0. Social comment	P. Personal needs (library, sharpening pencil, bathroom)
T. Off-Task (inappropriate exit)	W. Waiting
A. Secondary activity (reading a story not related to content, watching a film, playing a noncontent game, and so on)	F. Off-Task
	Passive
	Talking out
	Disruptive
	X. Interpersonal

Context Events (toggled keys in ELED1)

I. Individual
G. Group
C. Whole class

5. Instructional aids

NOTE: Adapted from Stallings, Needels, and Sparks (1987).

School Psychology

This section provides two category systems that have been used with success in a research and development capacity. The first is provided in summary form only as complete definitions and examples are available in the literature elsewhere. This first example is taken from positive social skill development work in public school settings and is representative of an all-inclusive and overlapping type of system intended to describe and analyze social interactions in peer activity contexts (refer to Sharpe, Brown, & Crider, 1995; Sharpe, Crider, Vyhlidal, & Brown, 1996, for detailed behavior and event definition examples and data-based outcomes).

The second example is presented in more detail and is based on the common use by school psychologists of some form of structured observations in the classroom when assessing a student or consulting about a student with a third party such as a parent (Rechsly & Wilson, 1996). Based on the appeal of applying generalized matching law theory (refer to Davison & McCarthy, 1988; McSweeney, Farmer, Dougan, & Whipple, 1986; Myerson & Hale, 1988, for a thorough discussion) and ecobehavioral theory (see Greenwood, Carta, & Atwater, 1991; Greenwood, Delquadri, Stanley, Terry, & Hall, 1985) to the description and analysis of behavior-environment interactions in ecological context, this second example developed and implemented by Shriver and Kramer (1997) provides insight into the complexity of interactions that a computer-driven observational effort may be capable of accurately describing and analyzing. The teacher codes were developed to assist in determining what level of behavior-coding specificity was required to examine possible interactions between teacher behavior and student behavior. Research efforts that used this coding system example found that more specific definitions of teacher behavior, and more specific delineations of similar behaviors, provided important information previously unrealized regarding teacher-student interactions.

Table 4 Tom Sharpe, Purdue University—Positive Social Behavior Category System Summary

Teacher	**Student**
Teacher observation	O. Organization
1. General observation	S. Sport engaged
2. Positive social observation	C. Cognitive
Organization directions	W. Waiting
3. Teacher directed	F. Off-Task
4. Pupil directed	
5. Teacher model	
6. Teacher exit	
Positive social response	
Skill content feedback	R. Following behavior
7. Verbal	L. Leadership behavior
8. Higher order	Issue resolution
9. Positive social feedback	A. Teacher assisted
0. Interpersonal	I. Teacher independent
Context events	
Grade level	Activity content
G1. First grade	Recorded as qualitative text
G3. Third grade	
G6. Sixth grade	

SOURCE: Taken From Sharpe, Brown, and Crider (1995).

Table 5 Mark Shriver, Meyer Rehabilitation Institute, University of Nebraska-Omaha Medical Center—Matching Law-Based Ecobehavioral Observation System Example

Task Structure	Student Behavior	Teacher Scheme
Opening	Reading aloud	Attentional context
Reading	Reading silently	Group
Related content activity	Writing	Target student
Closure	Listening	Peer student
	Transition	No attention
	Waiting	Off camera
	Verbal appropriate	
	Verbal inappropriate	**Behavioral context**
	Task appropriate	Instruction
	Task inappropriate	Listening
	Off camera	Approval
		Disapproval
		Business management
		Monitoring
		Other talk
		Independent work
		Off camera

SOURCE: Taken from Shriver and Kramer (1997).
Operational Definitions

Task Structure

Opening: Opening is coded for when the students are engaged in activities immediately upon entering the classroom. Such activities include removing and hanging up coats, setting up desks, pledge of allegiance, and so on. The students may have assignments that they are expected to be working on involving reading or writing. It is during this time that they give lunch orders, the teacher may take roll call, and the day's activities are planned and presented by the teacher to the students.

Reading: Reading is coded when the student's *primary task* is reading. The student may be reading silently or out loud. The student may be reading a reading primer or book, a library book, or any other material that only involves reading (not writing or workbooks). The student may be reading independently at his or her desk, in a small reading group, or with the whole class. In addition, if the student reads for the class or a group something he or she has written during the opening or the other reading activity, then that is coded as reading as well.

Related content activity: Related content activity is coded for any and all activities that occur during the class time that is primarily devoted to reading, but the student is engaging in activity that is not primarily reading. Such activity may include writing stories, drawing, working in a workbook or on a worksheet, and other paper-and-pencil tasks that may involve reading but the primary activity is writing. This category is also coded when the students are listening to the teacher lecture or discuss lessons or presentations related to reading. Basically, everything not associated with reading and opening is coded in this category.

Closure: Closure is coded for activities that occur at the end of the reading period and for activities that are *not* related to the learning and/or teaching of reading skills. Such activities include the transition to another academic period, cleaning up activities, getting and handing out snacks or materials, and so on. Closure is coded only when it is clear that the majority of the class is involved in closure activities.

Student Behavior

Reading aloud: Reading aloud is coded when the student is observed looking at reading materials and is saying aloud what is written in the printed material.

Reading silently: Reading silently is coded when the student is looking at reading material such as a book, primer, notebook, workbook or worksheet, and so on for at least 2 seconds *and* eye movements indicate that the student is scanning words, numbers, or letters. Rapid flipping of pages is coded task inappropriate and is not included in this category.

Writing: Writing is coded when the student is observed marking academic task materials such as a paper, ditto sheet, workbook pages, and so on with a pencil, pen, or crayon. Writing involves holding the writing instrument between thumb and forefinger and moving it in a manner likely to produce written numbers, letters, or words. Drawing pictures is not included in this category and is coded task appropriate or task inappropriate depending on whether the drawing is acceptable to the teacher's directions at that time.

Listening: Listening is coded when the student is looking at the teacher giving directions, commands, lecturing, or at another student who is asking or answering a question. Listening is also coded when two students are engaged in an activity together and the target student is looking at the peer or task as the peer talks. Listening is also coded when the target student is listening to another student or the teacher read aloud from a book or primer or worksheet. Even though the target student may appear to be following along reading silently, listening is still coded (and not reading silently).

Transition: Transition is coded when the student is required to change or get materials during and after an activity. For example, to fetch new materials or to put materials away, move to a new location in the room (e.g., desk to reading table), clean up, get into a line, and so on.

Waiting: Waiting is coded during those times when the student has completed an assignment or transition and is awaiting instruction or direction regarding what to do next. Examples include standing in line, waiting for students to quiet down while being quiet, sitting in desk and awaiting instructions, and so on. If instruction or direction has been given by the teacher and the student continues to wait and not follow that direction, then this behavior is coded as task inappropriate.

Verbal appropriate: This is coded when the student is observed verbalizing about his or her academic subject materials, teacher instruction, or other appropriate topics related to the lesson. The student may be directing his or her talk at the teacher, another student, or himself or herself. For example, Jim could be talking to Darlene and saying, "Which is larger on this page, seven or four?" Darlene answers, "Seven is larger, here, count seven fingers and then count four fingers." Jim then counts, "One, two, three, four . . ." To code verbal appropriate, talk does not have to be directly related to schoolwork but must be directly permissible and encouraged by the teacher and the specified task.

Verbal inappropriate: This category is coded when the student is observed talking aloud to a peer, teacher, or himself or herself about academic or nonacademic topics not related to the activity or task at hand *and* the student does not have teacher permission to be talking during that time. This category also includes verbal noises (e.g., grunts, laughing) that are not appropriate to the task or acceptable to the teacher.

Task appropriate: This category is coded when the student is engaged in play behaviors or other activity (e.g., working on the bulletin board, drawing, etc.) approved by the teacher. Task appropriate is coded whenever the student is behaving appropriately but the behavior cannot be coded in any of the other eight categories previous described above.

Task inappropriate: This category is coded when the student is clearly engaged in tasks that do not have the approval of the teacher and are not related to the present task at hand. Examples include working on academic or nonacademic tasks that are not currently assigned, avoiding the assigned task by coloring rather than working on a problem, passive behaviors such as sleeping or daydreaming, and disruptive behaviors such as hitting or acting out inappropriately.

Teacher Scheme/Attentional Context

Attending is generally defined as the teacher speaking to, or behaving toward, a student or group of students such that the teacher is directly aware of the student(s) behavior. Such teacher behaviors may include teaching, listening, general talking, watching a student, touching a student, and so on.

Group: Group is coded whenever the teacher is observed attending to the class as a whole or is attending to a group of at least three or more students (e.g., a reading group) of which the target student belongs.

Target student: Target student is coded when the teacher is observed to attend specifically to one target student. This may occur on a one-to-one basis or in the context of the larger class. A latter example includes teacher attending to the whole class during class instruction but asks a target student a question during class instruction. Attention includes both verbal and nonverbal attention (e.g., explaining a skill, observing a student perform a task, etc.).

Peer student: Peer student is coded when the teacher is observed specifically attending to another student (and not a targeted student). Examples include monitoring a peer paired with a target student in a reading activity.

No attention: This category is coded when the teacher is observed to not be attending to any student or group as defined by the previous three categories.

Teacher Scheme/Behavioral Context

Instruction: Instruction is coded whenever the teacher is actively instructing or giving a lesson to student(s). Teacher behaviors that may be coded include presenting information, asking questions, answering questions, writing on the blackboard, and so on, all as an explicit part of the current lesson's instruction (and not managerial assignments).

Listening: Listening is coded when the teacher is observed to be listening to a student(s)' presentation, reading aloud effort, recitation, questioning, and so on.

Approval: This category is coded when the teacher verbally or physically expresses approval to student(s). This may take the form of praise, encouragement, appreciation, or satisfaction with student work, conduct, or general class performance. This teacher behavior is generally associated with teacher attempts to increase a particular student behavior. Examples include teacher repeating of student answers with positive emphasis; gestures such as clapping, smiling, winking, waving, patting on the back; statements such as "I'm so pleased with your work," "good work, you can do it Jimmy," tangible rewards such as stickers and candy, and so on.

Disapproval: This category is coded when the teacher expresses disapproval in the form of dislike, dismay, dissatisfaction, or disgust with a student or group's academic or non-academic work, appearance, conduct, or performance. This category is generally associated with teacher attempts to decrease a student's behavior. Examples include gestures such as taking things away, shaking of one's head back and forth, frowning or grimacing; punitive measures such as placing a student in time-out; verbal statements such as "No, Jimmy, you never get those right," "What is wrong with you," and so on.

Business management: This category is coded when the teacher talks about class business, rules and regulations, daily schedules, the organization of future activities, and so on. All management activities such as handing out papers, collecting lunch money, taking roll, and so on are also included in this category. In addition, for business management to be coded, the teacher's attention must be directed at student(s). For example, if the teacher is looking for materials in the closet and is not attending to students, then this category is not recorded.

Monitoring: This category is coded when the teacher visually scans the larger class in a general observation way or is specifically observing a particular group of students or a particular student. Teacher observation for longer than two seconds must occur to record this category.

Other talk: This category is coded when the teacher talks to students or a particular student about information unrelated to academic or general classroom activities. This category includes all interpersonal interactions and includes examples of "I love your new dress," "How did you do in the soccer game last night?" and so on.

Independent work: This category is coded whenever the teacher engages in an activity that does not include any students in the classroom, and the students are not being attended to in any way. Examples include reading or writing at the teacher's desk, looking in a cupboard of materials with back turned toward students, and so on.

Off camera: A common code across all three subcategories, this category is coded whenever the target student or teacher (depending on the subcategory) is off camera for more than two seconds.

Special Education

Special education is one area within education that has provided a host of research and evaluation examples that use systematic observation methods and some form of behavior analysis. A first example, contributed by Jeff Sprague and his colleagues at the University of Oregon, provides an elegant illustration of an interbehavioral or sequentially based category system approach to the observation and evaluation of education settings in which moderate to severe mental delays are characteristic of the students served. This category system example provides insight into how one might first construct a set of response classes that describe in a more general way the interaction patterns that may occur between behavior-behavior and behavior-event relationships in the setting to be observed. In addition, this category system example provides a set of terms and definitions designed to document participant perceptions such as constructive versus coercive environment, in a behavioral way. For detailed examples of related research efforts, refer to Sprague and Horner (1992, 1994).

A second example contributed by Ken Simpson and his colleagues at Southern Illinois-Carbondale provides a composite illustration of how direct observational methods have contributed to evaluation practices in special education settings. A direct observation approach to research and evaluation in special education settings provides one means to federal public law mandates that speak to the necessity of monitoring the daily performance of students with disabilities. As Johnson et al. (1995) state, most observational tools developed for special education settings emphasize the systematic collection of measurable information, analysis of that information, and the direction and refinement of educational interventions based on those analyses. Observational and behavior analytic approaches to special education concerns have been proven effective in improving student academic performance (see Fuchs & Fuchs, 1986) and have documented a corresponding *lack* of student improvement in the absence of this type of information (Utley, Zigmond, & Strain, 1987). According to Croll (1986), an appealing method for assigning student and teacher actions to measurable categories provides the potential for very useful descriptions and analyses of participant interactions in therapeutic settings. The category system terms and definitions examples listed within Table 7 are drawn from a compilation of special education efforts (refer to Hunt, Alwell, Farron-Davis, & Goetz, 1996; Light, Collier, & Parnes, 1985a, 1985b, 1985c, for a more detailed discussion of term and definition development and use) and provide important summary illustrations of how category systems may be developed for special education evaluation purposes.

Table 6 Jeff Sprague, University of Oregon—Interaction Pattern Analysis Example Including Terms and Definitions

Student Behavior Codes		
Response Class	**Behavior Code**	**Definitions**
Problem behavior	Major	Those behaviors that result in pain, property damage, or major disruption and are judged by the teacher to be highly aversive.
	Minor	Those behaviors that do not result in pain, property damage, or major disruption. Minor problem behaviors will be rated by the teacher as mildly aversive.
	Noncompliance	Student does not engage in behavior that meets the requirements of the preceding teacher request for 10 seconds.
Compliance	With minor problem behavior	Student engages in minor problem behavior and complies with the current within their turn.
	Straight compliance	Student responds or attempts to respond to the current teacher request within 10 seconds of the request. Response must meet or approximate the requirements of the staff request.
Agent-action request	For attention	Student verbal statement or gesture that indicates a conventional attempt to access attention or interaction from the teacher (e.g., “look at me!”).
	For tangible	Student verbal statement or gesture that indicates a conventional attempt to access a desired item of food, toys, and so on (e.g., “I want a cookie,” which leads the teacher to an object).
Other or no response	No	No student response for 15 seconds following a teacher turn.
	Other	Any statement or gesture directed toward the teacher that does not meet the definition of any other student code.

Table 6 Continued

Response Class	Behavior Code	Definitions
	Teacher Behavior Codes	
Demand request	Initial request	A verbal statement by the teacher to the student to perform a behavior or task. An initial request only occurs when the behavior demands are not related to the previously occurring request (e.g., a new instructional trial with new demand content).
Maintain requests	Repeat or increase	A verbal statement or gesture by the teacher indicating a request to perform the same behavior or task stated in the current initial request. The request must indicate the same or increased expectations (e.g., behavior, criterion, and time frame).
	Correction	A statement or gesture by the teacher that indicates strong disapproval. This includes “no” statements, threatening to remove privileges, and raising voice volume.
Reduce or remove requests	Decrease	A secondary verbal statement or gesture directed to the student that is a modification of the initial request. This may include reducing the conditions or form of the request, providing prompts, reducing the criterion for completion, or the time frame.
	Reward/praise	Any statement or gesture by the teacher that indicates approval (e.g., praise, physical contact, smiles, positive statements, tangibles, etc.).
Other	Response	Any statement or gesture by the teacher that does not meet criteria for any other code.
	No response ignore	No response by the teacher following a student turn for at least 10 seconds.
	Response String Definitions	
Initial request or string start	(1)	A verbal statement by the teacher, directed to a student to perform a task, behavior, or activity. The statement must indicate the behavior to be performed or the outcome.
	(2)	The first student verbal statement or gesture that indicates a conventional attempt to access attention or interactions from the teacher (e.g., “look at me!”) after a response string end.

Table 6 Continued

	Teacher Behavior Codes		
	Response Class	**Behavior Code**	**Definitions**
		(3)	The first student verbal statement or gesture that indicates a conventional attempt to access a desired item of food, toys, and so on (e.g., "I want a cookie," leading the teacher to an object) after a response string end.
	Middle turns		All interactions that occur between an initial request and the response string end.
	Response string end	(1)	Compliance to teacher request (i.e., escape).
		(2)	Student obtains tangible (i.e., tangible).
		(3)	Teacher provides positive attention and student reduces or terminates problem behavior (i.e., attention).
		(4)	30 seconds or more of no turns by either the teacher or student.
		(5)	Teacher presents a request that differs in form or function from the preceding initial request.

Coding Interaction/Turn Intensity Hierarchy

Note. A final strategy to using this category system to code teacher and special education student behavior and behavior sequences (i.e., interactions or turns) is to code the most intense behavior performed during that turn.

Student Behaviors	**Teacher Behaviors**
1. Major problem behavior	1. Terminate string
2. Compliance with minor problem behavior	2. Repeat/increase request
3. Minor problem behavior	3. Decrease request
4. Tangible and attention requests	4. Reward
5. Noncompliance	5. Correction
6. Compliance	6. No response/ignore
7. Other response	7. Other response
8. No response	

Table 7 Ken Simpson, University of Southern Illinois-Carbondale—Composite Illustration of Special Education Evaluation Codes

Parameter 1: Structure	
Initiation:	Any verbal or active nonverbal behavior that engages or attempts to engage another person.
Acknowledgment:	Any verbal or nonverbal behavior that appears to be in response to an initiation.
Reciprocal interaction:	Exchanges that include both initiation and acknowledgment behavior.
Parameter 2: Function	
Request:	To ask for objects, actions, or information.
Protest:	To indicate a desire to avoid an undesired stimulus or to escape an ongoing stimulus.
Comment:	To make a remark or provide information.
Assistance:	To provide information or other assistance that helps the partner accomplish some outcome.
Greeting:	Any salutation to begin or end an interaction.
Parameter 3: Focus	
Social:	An interaction whose major purpose is the interaction itself.
Task related:	An interaction in which an outcome is accomplished that goes beyond social contact.
Parameter 4: Quality	
Mismatch:	A reciprocal interaction in which the quality of one partner's communication is positive and the other partner's communication is negative.
Neutral:	The reciprocal exchange is made with neither positive nor negative affect.
Affirming:	The interactive exchange is caring or loving.
Complimentary:	An interactive exchange in which one partner praises or compliments the other.
Sharing pleasure:	Partners are jointly participating in a pleasurable activity, or the reciprocal interaction itself is generating pleasure.
Humorous:	An interactive exchange in which one partner initiates a funny action or remark and the other partner responds appreciatively.
Displeased:	Both partners in the exchange express irritation, discontentment, indignation, and/or exasperation.
Angry:	Both partners in the exchange express hostility, resentment, and/or animosity.
Parameter 5: Communication Turns	
Presence:	Two or more participants involved in verbal or nonverbal communication.
Absence:	Two or more participants not involved in a verbal or nonverbal communication act where one is clearly expected.
Parameter 6: Communication Function	
Social convention:	A greeting among participants.
Request:	Asking for an object or set of materials, requesting additional information on a topic, requesting clarification on information received, or asking for attention from another participant.
Provision:	Providing information or clarification on a particular topic.
Imitation:	Imitation of a communication partner.

Table 7 Continued

Parameter 7: Mode of Communication	
Communication board:	Use of written signals to communicate.
Vocalization:	Any form of verbal interaction.
Nonverbal gesture:	Any form of nonverbal communication.
Parameter 8: Partner Identity	
Paraprofessional	
General education teacher	
Special education teacher	
Therapist/related services personnel	
Other adult	
Student without disabilities	
Student with disabilities	

Clinical Psychology

Clinical psychology is another area in which great benefit may be seen from a direct observational approach to the various interactions present within particular types of therapy sessions. The example contributed by Dennis Delprato and his colleagues in the Psychology Department at Eastern Michigan University provides insight into how a sequential or interbehavioral coding system in particular may provide valuable information related to effective and not-so-effective communication patterns with clients in group settings. Most clinical work to date of this type has originated within medical center communities working with therapy groups with individuals who exhibit a range of common problems from psychotic language, to substance abuse resulting in incarceration, to schizophrenia, to various other psychological challenges. Important to this example is the view that verbal behavior may be coded as a function of interactive and interconnected responses rather than independent and discrete behaviors. In this way, verbal interaction (or language) may be more appropriately viewed as a dynamic sequence of interconnected events rather than the inappropriate static view that a more traditional discrete behavior analysis might provide (refer to Bijou, Umbreit, Ghezzi, & Chao, 1986; Kantor, 1977; Ray, Upson, & Henderson, 1977, for a detailed discussion of this issue). Rather than stopping with a more traditional analysis of the number of times particular behaviors are emitted, a recommended analysis focus in this regard is on the probability of any verbal behavior occurring as a function of its time-based proximity with another behavior within the coding system. As other sections provide illustrations of operational definitions for behaviors and environmental events that are similar to those presented in this example, we offer only a summary of terms here.

Table 8 Dennis Delprato, Western Michigan University—Sequential Analysis Example of Therapy Interactions in a Clinical Setting

Therapist to Client
1. Statement of goals
2. Problem knowledge/description statements
3. Problem solution statements
4. Problem questioning
    - Group
    - Individual
5. Resistance confrontation
6. Compliance reinforcement
7. Client restatement
8. Interpersonal statements
9. Problem unrelated statements
10. Nonresponse to client statement

Client to Therapist
1. Nonresponse to therapist statement
2. Resistance/conflict statement
3. Defensive reaction
    - Verbal
    - Nonverbal
4. Guarded reaction
    - Verbal
    - Nonverbal
5. Statement of receptivity to therapist statement
6. Problem knowledge answer - inappropriate
7. Problem knowledge answer - appropriate
8. Unrelated comments

Client to Client
1. Problem related - unsupportive
2. Problem related - supportive
3. Problem unrelated - unsupportive
4. Problem unrelated - supportive

Context
1. Group therapy
2. Individual therapy

Ethology

Ethology is an area in the social sciences that has provided some pioneering work in the design and implementation of observational category systems and, in particular, some pioneering work in interbehavioral and field systems theory in relationship to those observational efforts. Roger Ray of Rollins College is one such pioneer on both of these accounts (see Ray, 1983; Ray & Delprato, 1989, for a detailed introduction to this literature). This section provides one illustration of how a set of terms and definitions may be constructed to observe animal interactions in controlled settings (refer to Astley et al., 1991). According to Astley et al. (1991), "the ability to record and analyze ongoing social behavior among animals depends on the development of adequate measurement systems that lend themselves to mathematical analysis . . . the most commonly used approach is formal coding" (p. 173). The primary purpose of the coding system example included in this section was to provide a data collection tool for enhancing understanding of social behavior in animals and to begin to understand the cardiovascular responses associated with those particular behaviors. By correlating cardiovascular response with behavior, this example also attempts to overcome what some have criticized (refer to Johnston & Pennypacker, 1980) as the arbitrary nature of breaking up the natural time-based stream of behavior-behavior and behavior-environment interactions. As complete operational definitions, and the rationale for constructing those definitions, are laid out in explicit detail within Astley et al., we provide only a summary of terms contained within the coding system here. This illustration also provides another example of how a category system might be constructed within an interbehavioral or field systems framework as articulated in our earlier chapters.

Table 9 Roger Ray, Rollins College—A Code to Integrate Animal Behavior and Cardiovascular Response

Locomotion and Posture	Behavior
1. Unidentifiable	1. Unidentifiable
2. Lie	2. Feeding
3. Sit	(a) Calm
4. Stand on two legs	(b) Active
5. Stand on three or four legs	(c) Excited
6. Walk on two legs	(d) Anxious
7. Run	3. Oral
8. Climb or jump	(a) Upright drinking
9. Hang	(b) Head down drinking
10. Lunge	(c) Licking object
11. Walk/stand sequence	(d) Biting object
	4. Grooming
Environmental Setting	(a) Groomed
	(b) Groomed while orienting to environment
1. Entrance opened	(c) Presents for grooming
2. Food brought into environment	(d) Grooms another
3. Food delivered	5. Maternal
4. Food removed from environment	6. Sexual
5. Threat by other	(a) approach female
6. Exit opened	(b) Mount with thrusting
	(c) Copulation
	(d) Toward adolescent
	(e) Approached by male
	(f) Present hindquarters
	7. Submission
	(a) Low level
	(b) Mid level
	(c) High level, retreat from animal
	(d) High level, retreat from person
	(e) Captured
	(f) Present to opposite sex
	(g) Present to same sex
	8. Aggression
	(a) Low level
	(b) Mid level
	(c) High level, attack animal
	(d) High level, attack person
	(e) Fight
	(f) Hindquarters, same sex
	9. Idle/nonspecific
	(a) Baseline activity
	(b) Extremely active, irritated, upset
	10. Other
	(a) Huddling
	(b) Play
	(c) Other

SOURCE: Taken from Astley et al. (1991).

Chapter 4: Reliability, Staff Training, and Future Directions

> [The issue of reliability] confronts us with exactly what we are doing, as opposed to what we have come to imagine we are doing. It is a telling fact that teachers and researchers are surprised by some of the things they find in [their own] descriptive accounts.
>
> Lawrence F. Locke

Once familiar with how to use the software tools described in the technical guide, and having decided on a set of terms and definitions to use as an observational category system (using the procedures detailed in the Chapter 3 pdf file), an individual or group of professionals needs to next make sure that they are consistently accurate collectors of the type of information they are about to collect and analyze. Regarding this next issue, we have attempted through the development of more capable computer-based tools to provide a set of procedures more amenable to the traditionally time- and labor-intensive task of assessing data reliability.

Somewhat related to the issue of ensuring the accurate and reliable collection of behavior-environment data, it is also important to determine procedurally just how the information of interest should be collected. Regarding this latter issue, various types of well-documented recording tactics merit discussion in light of the enhanced recording capabilities that our software tools provide.

If we are going to place an emphasis on the quantitative counting of observable events in our data collection tasks, then it becomes important to provide a set of procedures to make sure that when we are counting things, we get it right. In other words, we want to make sure that the data record that describes an experimental or evaluation setting of interest is as accurate a reflection of what actually occurred as possible. To do this, a summary of accepted (and recommended) data reliability assessment methods is provided, and a summary of traditionally used data recording tactics is provided in the context of recommending a *real time* approach.

A computer-based approach to the observation and analysis of behaviors and events also provides many research and development possibilities not previously attainable without the aid of these tools. This is particularly true in light of how rapidly computer technologies are advancing and how quickly the capabilities of these technologies improve in levels of sophistication. In this regard, the last portion of this chapter summarizes some appealing research, evaluation, and teaching directions that may be realized through the use of the software tools provided with these materials and through the use of similarly developing technologies.

A Recommended Three-Step Reliability Process

For a summary discussion of reliability and a set of recommended procedures to use to ensure adequate reliability of data gathering practices, it is first important to distinguish between what is meant in the behavioral literature by the terms *observer agreement*, *reliability*, and *accuracy*. Generally, observer agreement procedures provide an indication of how closely two or more independent observers of the same behavior and event occurrences agree in their data recording when they are involved in collecting data on the presence or absence of those same occurrences. According to Johnson and Bolstad (1973), the terms observer agreement and reliability have been used interchangeably in much of the behavior analytic literature, though it remains important to understand the distinction between the two in a conceptual sense. The distinction is as follows: while observer agreement is the extent to which two or more individuals agree on the presence or absence of behavior and event occurrences, reliability is the consistency with which behavior and event occurrences will be recorded in the same way when they occur in the same way at different points in time. In other words, while two different observers may agree with one another in terms of their data records, those data records may not necessarily be reliable if each individual records in an inconsistent manner over time.

While reliability refers to the consistency over time of data recording efforts, the last term of accuracy refers to whether a particular data record is representative of behavior and event occurrences as they actually occur within an observational setting of interest. In other words, a data record is said to be accurate if it reflects what actually happened in the observational setting. Similar to the relationship between observer agreement and reliability, two independent observers may be reliably (and consistently) collecting data on the same observational settings, but that does not necessarily mean that their data records are accurately reflecting what is actually occurring in those observational settings (i.e., both observers could conceivably be reliably and consistently wrong in their observations).

Given the definitions of observer agreement, reliability, and accuracy, the ultimate challenge of a scientific enterprise devoted primarily to the recording of observable events is to try to ensure that the data collected to describe a particular setting is indeed accurate. Behavior analysis research efforts most often advocate the use of various agreement tests and related reliability procedures in an effort to increase the *probability* that the data collected for a particular research or evaluation purpose is accurate. While data accuracy cannot be *ensured* in a pure sense through the use of these procedures, their use is designed to significantly increase the probability of accurate data recording efforts.

Kazdin (1982) provides three important rationales for including procedures to ensure agreement in any research or evaluation enterprise involving the counting of behaviors and events. First, such procedures ensure a degree of consistency in the recording of similar types of information. Second, if agreement is ensured across data records, the potential for bias and changes in behavior and event occurrence interpretations across observers is minimized. Third, implementing agreement procedures provides a check for whether the behaviors and events to be observed are well defined. Conducting agreement checks among observers during the course of an experiment or an evaluation activity is one way to ensure that different observers are interpreting behavior and event definitions and recording the occurrences of those behaviors and events in the same way. In other words, agreement checks help determine if the terms and definitions provided within a particular category system meet the objective, clear, and complete definitions provided in the Chapter 3 pdf file.

Step 1: Developing a Criterion Standard

We advocate three general steps to be included in any set of procedures designed to ensure the reliability and, it is hoped, the resultant accuracy of data records. First, we recommend the development of what we term a *criterion standard*. A criterion standard is simply a videotape record of similar situations as that of the experimental or evaluation setting to be observed for data recording purposes. For example, a teacher educator who is interested in using a category system approach to the evaluation and feedback+goal-setting experiences within practice teaching situations to be observed would first make a videotape of similar practice teaching situations. When making the videotape record, that teacher educator also would make sure that all of the behaviors and events contained within the category system to be used for evaluation purposes occurred with some frequency throughout that videotape record. Taking another example, if an ethologist was interested in studying the various interaction patterns of a particular species of baboons in the wild (no correlation with our first example intended), that ethologist would first make a videotape record of similar baboons in similar settings to the one to be observed making sure that all behaviors and events of interest occurred frequently on that videotape record.

Once a videotape record containing examples of the behaviors and events to be observed is complete, a next step involves dividing up that videotape record into equal time segments and recording using a video-player time counter the start and stop times of each segment. To illustrate, if the teacher educator in our example prepared a 45-minute videotape record, he or she might next break up that record into nine, 5-minute segments and record using the video player's time counter the start and stop times of each segment.

Once a videotape record of behavior and event examples is made and segmented according to time units, two lead researchers (or evaluators as the case may be) need to next synchronize a data collection mechanism to the start times of each time-based segment on the videotape and collect a data record for each segment. If using the software contained with these materials, this simply involves starting the data collection application at the same point in time as the starting of the five-minute segment of interest on a videotape record (e.g., unpausing the videotape player at the same time the data collection application is turned on). Taking our teacher educator illustration, each of the two lead observers would conduct an independent recording of each of the nine five-minute segments, with each investigator conducting a respective and independent five-minute data recording episode on the same day. Continuing with our teacher educator example, once two lead teacher educators have completed the recording of all nine five-minute segments on the videotape record, the data records from each respective time segment should be compared across teacher educators using the reliability application contained within the software provided.[1] This step should be repeated until the two teacher educators agree to a minimum level of agreement (typically .80 to .85; refer to Kazdin, 1982; Bakeman & Gottman, 1986, for a more thorough discussion of minimum levels of acceptable agreement) across each of the nine five-minute segment comparisons.

Once two lead researchers (or evaluators) consistently come to agreement above a minimum specified level for all of the videotape record segments, the data recording and comparing process just described should be undertaken again after a period of two to three weeks has elapsed. If the two lead researchers still consistently agree across all videotape record segments above a minimum specified level, then a final comparison step is conducted. This final step involves the comparison of each observer's first set of data records to each of their second matched set of data records. Using our teacher educator illustration, the first teacher educator would compare the first five-minute data record of their respective first set of nine to their first five-minute data record of their second set of nine, and the second five-minute data record to the second data record of the second set of nine, and so on until nine comparisons were made. The second teacher educator would do the same. If each teacher educator agreed with himself or herself to a minimum specified agreement level for all nine comparisons across time, then the development of a criterion standard is complete. At this point, a set of data records has been generated that corresponds to specified time-based segments of a videotape record, and the data records have been determined consistently reliable across two independent observers and have been determined consistently reliable over time within observers

Observer #1:	a1 a2	b1 b2	c1 c2	d1 d2
Observer #2:	e1 e2	f1 f2	g1 g2	h1 h2.

According to the illustration above, and continuing with our teacher educator example, a criterion standard is successfully produced if two teacher educators agree to a minimum specified level for comparisons a1-e1, b1-f1, and so on, throughout the number of observational segments (commonly termed *interobserver agreement*) *and* agree to a minimum specified level for comparisons a1-a2, b1-b2, e1-e2, f1-f2, and so on, throughout the first and second observations of the same segment for each respective observer (commonly termed *intraobserver agreement*). Again, within all of these illustrated comparisons, agreement must be consistently above the minimum specified level across all of the time-based segments contained on the videotape record.

If, however, the two lead researchers do not consistently agree with themselves (i.e., the second intraobserver agreement step), then the first interobserver comparison step must be undertaken again until the two lead observers return to consistent agreement with one another to a minimum specified level across all time-based segments. What we have found helpful in this instance, as this happens more often than not when initially developing a criterion standard, is for the two lead observers to slowly go through the videotape record together and talk over various disagreements and challenging behavior or event determinations. After this type of discussion, the two lead observers should make another effort to come to consistent and independent agreement with one another over all of the time-based segments contained on the videotape.

Another less likely occurrence during this process is that two lead observers will consistently agree with one another (comparisons a-e, b-f, c-g, etc.); however, they will not consistently agree with themselves over time (comparisons a1-a2, b1-b2, etc.). This problem is commonly referred to as *observer drift* and is a challenge important to check for to ensure that it is avoided, both in the development of a criterion standard and during the collection of actual data records for a particular experiment or evaluation project. If observer drift is occurring, then changes in interpretation of the behavior and event definitions are occurring with particular observers over time and will result in changes in the data records due to observer bias and interpretation rather than changes in the actual occurrences of the behaviors and events being observed and recorded. If this is the case during the development of a criterion standard, then we recommend that the lead observers first go through the entire videotape record together and discuss challenging behavior and event determinations and discuss the instances in which observers have been changing their recording practices. After this type of discussion, a more rigorous agreement comparison procedure should be implemented to include (a) interobserver a-e, b-f, and so on comparisons; (b) intraobserver a1-a2, b1-b2, and so on comparisons; and the addition of (c) interobserver a1-e2, b2-f1, and so on comparisons across time. Using our teacher educator illustration, these three forms of comparisons should be undertaken according to the steps discussed until both lead observers have consistently agreed across all nine five-minute videotape segments at the same point in time, at another later same point in time, and with themselves and with each other across different points in time. If all three forms of comparisons are undertaken, and a minimum level of agreement is consistently reached within all three forms of comparison, then we can stipulate with a degree of confidence that the final data records that represent each time segment contained on the videotape record are accurate representations.

Though our discussion of the development of a criterion standard may seem daunting in terms of the time and effort invested, we feel that it is an important step to increasing the probability of accurate data records once an experimental or evaluation project is under way. Successful and rigorous completion of this criterion standard development step is crucial to ensuring the reliability and hoped for accuracy of data records during an experimental or evaluation project. This is also a step most often overlooked when procedurally discussing and reporting behavior analytic data in the education and social science literatures. The importance of this step will become more apparent once involved in our second recommended general step of staff training to the criterion standard that has been developed.

Step 2: Staff Training to Criterion

Once a criterion standard has been developed for a particular set of research or evaluation projects and includes a time-segmented videotape record and a set of data records that correspond to those time segments, a next general step is to train a group of observers to conduct reliable, and hopefully accurate, data collection efforts. To do this, we recommend as a first step that the lead investigators who prepared the criterion standard first hold an informal discussion with all potential observers over the terms and definitions contained within the category system to be used and use the criterion standard to show examples of what behavior and event occurrences tend to look like. Next, education on the use and application of the software contained within these materials is necessary until the observers to be trained are familiar with how to successfully operate the data collection and data analysis applications, the reliability application in particular.

Next, using the reliability software application contained with these materials, we recommend that all observers be trained to observe the segments specified within the videotape record contained within the criterion standard materials for the purposes of collecting data records. Using our teacher education example, this would involve a particular observer synchronizing the videotape record with the data collection apparatus and collecting data on the first five-minute segment of the videotape record. Once accomplished, this observer would next tag both the first five-minute criterion standard data record and the data record they just collected for the purposes of conducting a reliability assessment with the software application. After an observer recorded and compared his or her data collection effort with the respective data record contained with the criterion standard materials, that observer should repeat this collection and comparison procedure with the second five-minute videotape record segment. After three consecutive time-segment attempts on a particular day, if an observer meets a minimum specified level of agreement when comparing his or her data records to the records of the criterion standard on all three attempts, then that observer is considered successfully trained. If a minimum specified level of agreement is not met on any one of the three attempts, then we recommend that that observer stop any further data collection attempts until the next day. In essence, we recommend that these three data collection attempts on consecutive time segments of the criterion videotape record be repeated until an observer successfully meets a minimum specified agreement level on three consecutive five-minute segment collection and comparison attempts.

Once successfully trained according to the specifications we have discussed, and assuming a certain level of commitment and enthusiasm for the observation tasks to be conducted, lead observers should be confident in the initial reliability and potential accuracy of data collected by these trained observers. For large groups of professionals who plan to use the same category system for a similar group of observational tasks from year to year, we additionally recommend that staff training be conducted for all observers on a yearly basis to ensure that all individuals within a group remain successfully trained. Of additional appeal to the time-consuming task of our initial recommended step of preparing a criterion standard is that once prepared, this same criterion standard may be used to train a host of observers over a number of years. In other words, once a criterion standard has been successfully prepared, many observers may be trained for many research and evaluation projects in a relatively time-efficient manner, as long as each data collection task uses the same category system and specified research and evaluation tasks are focused on the observation of the same behavior and event occurrences.

Step 3: Implementing Interobserver Checks

Once a criterion standard has been prepared and a staff of observers has been successfully trained, one remaining step to ensure the reliability and hopeful accuracy of data records is recommended. This step involves two trained observers independently collecting a data record on the same scheduled observational episode. It remains as Kazdin (1982) states that while there are no definitive rules for the number and timing of interobserver checks, it is generally recommended that such checks take place at least once or twice per experimental phase of an experiment, or if conducting observations for evaluation purposes, at least once per month over the course of the experience. A few considerations are also important in relationship to how often interobserver checks should be conducted. These include the number of observers involved in a project, the complexity of the observation system used in terms of number of categories, and how well two independent observers are agreeing with each other during a particular observation episode. If, for example, only a few observers are involved in using a fairly simple category system and they agree with one another at a high level with the first interobserver check, fewer interobserver checks would be necessary in the future. If, on the other hand, a large staff is involved in using a very complex observation system and is experiencing difficulty reaching a minimum specified interobserver agreement level, then interobserver agreement checks (and potentially additional staff training) should occur more frequently.

In practice, to conduct interobserver checks, two trained observers would first independently generate a data record from the same scheduled observation. Once accomplished, the data records of the two evaluators would be compared to provide a level of agreement. If agreement falls below a minimum specified level, then those observers should be retrained according to the staff training procedures we have recommended to bring one or both of these observers back to a minimum accepted level of agreement. After retraining, another interobserver check should be conducted with these same two observers to determine if their agreement levels have improved.

Reliability Formula Summary

There are a number of ways, statistically, that agreement between two observational data records may be calculated. Many of these methods have frequented the education and social science research literatures. In addition, many behavior analysis texts provide a detailed summary of the mathematical equations and the relative benefits and drawbacks of each method when calculating levels of agreement. We therefore provide only a brief summary of the statistical formulae that are contained in the software tools with these materials. For a more detailed discussion, refer to Bakeman and Gottman (1986), Cooper et al. (1987), Kazdin (1982), or any number of behavior analysis textbooks.

Frequency

This method is one of the earliest represented forms of conducting agreement checks and is perhaps one of the easiest to perform. While it is most often used when only a frequency or simple event count is used as a recording measure, it can also be used with intervals or duration of behavior if those time measurements are segmented into equally specified units amenable to counting. The *frequency* formula is as follows:

$$\frac{\text{Smaller Count Total}}{\text{Larger Count Total}} \times 100 = \text{ \% Agreement.}$$

In essence, when two observers' data records are compared, the smaller of a behavior or event count from one observer is divided by the larger behavior or event count from the other observer. This number is then multiplied by 100 to provide a percentage. When using the simple frequency method of assessing agreement levels contained in our software applications, a percentage of agreement is given for each event contained within the two data records according to the above formula and according to the frequency counts for each behavior and event contained within each of the two data records.

While this method of determining levels of agreement between data records is useful and appealing in its simplicity, and can provide a quick check to see if agreement falls in a relatively high or low percentage level, it has one major drawback. Comparing totals of frequency counts does not provide information related to agreement on the actual instances of each behavior or event as it actually occurred in time. For example, one observer may inappropriately record many noninstances of a particular behavior and not record some occurrences. The second observer may record occurrences of the same behavior accurately. In this case, one observer is recording accurately and the other is not. While only one of the data records is accurate, each may have a very similar or identical total number of behavior occurrences and, hence, will be assessed as having a high level of agreement between the two when using the frequency agreement method for comparison.

Point-by-Point

The second statistical formula provided in our software applications is perhaps the most commonly accepted and used formula to date in the literature and is termed *point-by-point*. Again operating on a frequency or discretely segmented measure, this approach is an attempt to take into account whether there is agreement on each recorded occurrence of particular behaviors or events. The point-by-point formula consists of the following:

$$\frac{\text{Number of Agreements}}{\text{Number of Agreements} + \text{Number of Disagreements}} \times 100 = \%\ \text{Agreement.}$$

When using our software application, this formula first takes the number of agreements for which each observer agreed occurred for each behavior and event. This is conducted by using the smaller of the two agreement totals for a particular behavior or event as the number of agreements on that behavior or event. Next, this formula counts the number of times each observer disagreed on the occurrence of a behavior or event. This is conducted by taking the larger agreement total for one observer and subtracting from it the smaller agreement total from the other observer, for each behavior and event. The agreement number and the disagreement number is then plugged into the formula above and calculated. The outcome is a percentage of agreement and is calculated for each separate behavior and event contained within each data record compared.

The point-by-point method of determining levels of agreement provides a distinct advantage over a simple frequency method as it allows a determination of agreement taking into account behavior and event occurrences and not simply totals. Two potential drawbacks to this method exist, however. First, agreement on nonoccurrences of behaviors and events is not included in the method, as inclusion of nonoccurrence information provides a danger of artificially inflating the agreement percentage. In other words, observers may more readily agree on nonoccurrences, and absences of a behavior or event may be more prevalent than occurrences in an observational episode. What is assumed most important to assessing agreement is the accurate recording of the occurrence of a behavior or event. If nonoccurrences were included in an agreement analysis, then a large number of nonoccurrences in relation to occurrences may mask the ability of observers to reliably and accurately make challenging discriminations of particular behavior and event occurrences.

A second potential drawback to the point-by-point method is that when behavior and event occurrences are high, agreement levels are typically artificially high. In other words, a certain level of agreement between observers will be calculated with this formula due simply to chance or random recording.

Cohen's Kappa

Though often criticized due to their inherent complexity, some more sophisticated statistical/correlational methods are available that have been developed in an attempt to more capably control for nonoccurrence and chance agreement when comparing data records (see Issues in Interobserver Reliability articles in the *Journal of Applied Behavior Analysis*, *10*(2), 1977, for a more detailed discussion). Though little used and reported in the experimental literature in the education and social sciences, we provide Cohen's (1965) kappa in our software package as (a) this method is most often recommended when searching for a more sophisticated agreement statistic, and (b) through the use of computer-based technology, more sophisticated methods of assessing agreement become much more time and labor efficient. Cohen's kappa is conducted using the following formulae:

$$\frac{\text{P-Agreements} - \text{P-Chance Agreements}}{1 - \text{P-Chance Agreements}} = \underset{(0.00-1.00)}{\%\ \text{Agreement}}.$$

In this first formula that scribes completely the kappa measure, P-Agreements is calculated using the following:

$$\frac{\text{Number of Agreements on Occurrences} + \text{Number of Agreements on Nonoccurrences}}{\text{Total Number of Agreements} + \text{Total Number of Disagreements}}.$$

The P-Agreement formula is essentially the same as the point-by-point formula with the exception that it additionally takes into account agreements and disagreements on nonoccurrences of behaviors and events.

In the kappa formula, P-Chance Agreements is calculated using the following:

$$\frac{\left(\begin{matrix}\text{Number of} \\ \text{Occurrences} \\ \text{Observer 1}\end{matrix} \times \begin{matrix}\text{Number of} \\ \text{Occurrences} \\ \text{Observer 2}\end{matrix}\right) + \left(\begin{matrix}\text{Number of} \\ \text{Nonoccurrences} \\ \text{Observer 1}\end{matrix} \times \begin{matrix}\text{Number of} \\ \text{Nonoccurrences} \\ \text{Observer 2}\end{matrix}\right)}{\text{Total Number of Recording Intervals Squared}}$$

The specified advantages of using this formula (or those like it) is that it, in theory, takes into account agreement on nonoccurrences of behaviors and events and that it corrects for agreement due to random chance contained in observational recording episodes. A major drawback to this method of assessing agreement is that it is limited to the recording of the presence or absence of single behaviors within a specified time interval. In other words, to use this method with success, the method of observation must include segmenting the total observation period into equally spaced time segments and then limiting the observation process to the recording of the presence or absence of a single behavior within each specified time-segment. If multiple behaviors are to be recorded within the same specified time-segments, then a separate agreement analysis on each behavior or event of recording interest must be conducted if choosing an agreement measure such as Cohen's kappa. When recording the presence (or absence) of multiple behaviors and events in real time as they actually take place, this method of determining agreement between data records is considered by many to be far too time consuming and to potentially provide inappropriate agreement quotients.

While we have incorporated one mean of assessing agreement in our software tools that potentially takes into account nonoccurrence and chance issues, the matter of providing a formula that definitively corrects for these artifacts and does not introduce new statistical concerns to the assessment method remains, as Kazdin (1982) suggested more than 15 years ago, a topic of statistical debate. While many complex statistical variations now exist in the literature in attempts to overcome nonoccurrence and chance issues, most experimental and evaluation projects interested in measuring agreement continue to rely on the point-by-point method until a definitive alternative is provided.

Recording Tactics

Once category systems have been developed and a staff is prepared for the collection of reliable and accurate data records, a final decision must be made prior to actual data collection. This decision relates to the choosing of a particular recording tactic and related data measure. In other words, most traditional quantitative observation methods require the choosing of a set of procedures for actual collection of the data record and the basing of those procedures on the type of measure a researcher or evaluator is most interested in obtaining. Given the significantly enhanced capabilities of computer-based recording tools like those contained with these materials, however, these decisions do not have as much impact on the possible data outcomes of a particular research or evaluation project. In this regard, we summarize some of the more widely accepted recording tactics seen in the traditional behavior analysis and systematic observation literatures, all in the context of recommending a recording tactic that was not feasible prior to the advent of more capable computer-based recording tools.

Some Traditional Approaches

Most behavior analysis or systematic observation methods texts provide a detailed procedural description of recording tactics popularized by researchers and evaluators alike. The most popular of these traditional tactics fall into two general areas, one specifying the type of measurement an observer is interested in and one specifying how the data may be collected in terms of the timing of observing and recording that data. Two popular tactics specifying measurement type are typically termed *event recording* and *duration recording.* Tactics specifying timing of observations are typically termed *time sampling* and include the most frequently used *whole-interval*, *partial-interval*, and *momentary time-sampling* techniques. Each of these tactics has typically entailed the use of a coding sheet, a stopwatch, and a pencil, in which behaviors and events are recorded by hand and in which multiple personnel are often necessary to perform various recording tasks. Until recently, choosing within these tactics has been thought to be fundamental in any recording procedure that counts the presence or absence of behavior and environmental events in some way. Traditionally, it has been important to detail the form and character of these tactics in relation to the various benefits and challenges that each tactic provides to the reliable and accurate recording of certain types of quantitative information. In most respects, the choosing of a recording tactic has been thought critical to the unit of measurement that a researcher or evaluator is interested in and critical to the preferred way that data are to be reported.

We argue that while researchers and evaluators need to remain sensitive to these recording tactics and to their various advantages and limitations, the necessity of matching a particular recording tactic to a particular observational measure or data reporting method need no longer be the case when using more capable computer-based tools to record and analyze observational data. This is primarily due to the more capable and more inclusive nature of computer-based data recording methods. Nevertheless, we provide a summary of each of these more traditional, pre-computer-based recording tactics to provide a foundation on which to recommend a more contemporary and more inclusive tactical alternative: recording in *real time.*

Event recording. The first of these traditional recording tactics, event recording, focuses on the counting of the frequency or number of behavior or event occurrences for a particular observational episode. With this method, focus is only on counting the number of times each behavior and event in the category system occurs. This tactic is recommended in situations in which the behaviors and events of interest occur with relatively high frequencies and relatively short duration. If each observational session within a particular experiment or evaluation enterprise consists of about the same length of time, then the most appropriate measurement choice in reporting event recording data is a simple numerical count. If observation sessions within a particular experiment vary in terms of length of time, then using a rate (i.e., average number per minute or other unit of time) measure to report data is a more appropriate choice. From the event recording of numerical counts, ratio data among particular behaviors and events, percentage data of one behavior or event relative to others, and celeration data (i.e., the relative amount of increase or decrease in behavior or event occurrences from observation period to observation period) may also be calculated and reported.

Duration recording. This second traditional recording tactic focuses on the recording of how long particular behaviors and events tend to occur within particular observational episodes. With this method, the individual duration of particular behaviors and events is recorded and information such as the longest and shortest duration, average duration and ranges or standard deviations of duration, and length of time between behavior and event occurrences is the primary measurement focus. This tactic is recommended in situations in which behaviors and events occur infrequently and each occurrence tends to last for a long period of time relative to total observational time. Percentages of total observational time taken up by the occurrences of each behavior and event and ratios among those percentages may also be reported with this method.

Time sampling. This set of tactics provides various procedures in which the presence or absence of particular behaviors and events is recorded during specified time intervals. These tactics require a decision as to when the data should be collected. With all time-sampling methods, each entire observational episode is first broken down into logically specified time segments. Once segmented, an observer would then record the occurrence (or nonoccurrence) of behaviors and events of interest within each logically specified segment according to the tactics specified within a particular procedure. With whole-interval time sampling, an observer records within each specified time segment whether a specified behavior or event occurred for the entire time segment or interval period. With partial-interval time sampling, an observer records if a particular behavior or event occurred within each specified time segment and may also record the number of behavior occurrences or the percentage of that time segment in which a behavior or event occurred. Momentary time sampling provides a variation from these first two tactics. In momentary time sampling, an observer records the occurrence or absence of a particular behavior or event only if it is occurring or absent at the precise point in time at which each specified time segment ends.

An education example may serve to illustrate these three time-sampling, or interval recording, tactics. If a teacher educator was interested in evaluating a particular instructional episode, for example, he or she might divide up a particular special education resource setting to be observed into logically specified 2-minute periods over the course of a 40-minute instructional session. If interest was in the behavior of appropriate student responding, the observer would record (a) whether appropriate responding occurred at all during the course of each specified time segment if using a whole-interval method, (b) the number of appropriate responses that were observed within each 2-minute segment if using a partial-interval method, and (c) record if the student was engaged in responding at the end of each 2-minute segment if using a momentary time-sampling method. These methods also provide two additional measures of number of intervals and percentage of intervals in which a behavior or event simply occurred or occurred at a desired level of occurrence. All of these time-sampling, or interval-recording, tactics have typically been reserved for observations in which one or only a very few behaviors and events are to be recorded and reserved for observations in which time segments are typically connected with the presence or absence of a treatment or intervention designed to change behaviors and events of interest in some productive way. As the number of behaviors and events to be recorded increase, and the rapidity of potential behavior and event occurrences increases, this method becomes less appropriate.

Time-sampling procedures have largely been derived due to the inherent challenge of recording information on multiple behaviors and events throughout an entire observational session. With each of these methods, time is provided between each logically specified time segment for data recording purposes. During these second time segments, or down time, data are being recorded by hand and observations are not taking place. For example, an observer who is interested in recording the number of disruptive or self-injurious episodes of a particular special education student mainstreamed into a regular elementary education classroom would observe for the occurrence of those behaviors during each observational time segment and would then record the occurrence or number of occurrences of those behaviors during each period of time devoted to the recording of information. Time-sampling methods provide an important way to determine if certain behaviors or events are occurring at certain specified points in time. One of the greatest drawbacks to these tactics, however, is that when observing behaviors and events that occur frequently and are of short duration when they occur, an inaccurate representation of an experimental or evaluation setting may be derived due to the loss of important information during each time segment that is devoted to recording behaviors and events. In other words, valuable information may be lost when devoting portions of an observational task to recording data rather than actual observation.

Recording in Real Time

Though the above recording tactics have provided a wealth of useful and important information to various education and social science concerns, we remain convinced that there is much more to be learned about interactive settings through more inclusive means to observing behavior-behavior and behavior-environment interactions. We also feel that more capable computer-based tools, when used according to a field systems or interbehavioral theoretical perspective to the observation and analysis of interactive settings (refer to Chapters 1 and 2), provide two important additions to observational efforts. First, more complete and inclusive description and analysis are made available. As one becomes more familiar with the software applications provided with these materials, it is clear that computer-based tools will enable an observer to collect information on many more behaviors and events at the same time and will enable the provision of all of the measures listed below in relationship to the many traditional recording tactics we have discussed:

Event Recording: Numerical counts, rates, ratios, and percentages among behavior counts.

Duration Recording: Average length of occurrence, range, and standard deviations of occurrence lengths; shortest and longest length of occurrence; percentage of total observational time; length of time between behavior and event occurrences.

Time Sampling: Number and percentage of time segments in which a target behavior or event occurred, number and percentage of time segments in which a target behavior or event occurred at a minimum specified level.

Second, using computer-based data collection and analysis tools according to a field systems or interbehavioral framework will enable observers to look at and analyze their data in new and innovative ways. Many times, for example, it may be necessary to examine several different measures (e.g., number, rate, percentage, ratio relationships, etc.) of the same behavior or event to come to a more complete understanding of the dynamics of a setting in which a complex variety of behavior-behavior and behavior-environment interactions are rapidly occurring. We feel strongly that a multidimensional examination of the many individual characteristics of each behavior and event is important to a better understanding of how most education and social science settings operate. In addition, we feel that the provision of an analysis of the patterns in time among behaviors and events as they actually occur temporally, or over time, is important to an explicit description and understanding of how certain behaviors and events tend to interact (see Hawkins, Sharpe, & Ray, 1994; Odom & Haring, 1994; Sharpe, 1997a; Sharpe, Hawkins, & Lounsbery, 1998, for a more thorough discussion of these two issues). In other words, we feel that it is important to describe and analyze the potential impact that certain behaviors and events have on certain others by describing and analyzing their relative proximity to one another in terms of their time of occurrence in a particular setting.

Most traditional and non-computer-based behavior analysis and systematic observation methods have had a primary requirement that behaviors and events be defined in a discrete way with definitive start and stop times of occurrence. The term *discrete* simply means that a definitive beginning and end time to each behavior and event being recorded be defined so as to be clear to those observing the occurrences of those events. The three major constraints of most earlier observational work in this vein have been that (a) we have had the capability to observe only a few behaviors or events at a time and have been forced to view all others operating within a setting of interest as external (or extraneous, see Kerlinger, 1986), (b) we have not had the tools to look at multiple behaviors and events in terms of explicit description and analysis of their time-based or temporal association with one another as they actually occur, and (c) we have often been forced to segment total observation periods into intervals with portions of the total observation period devoted to recording data rather than actual observation between these specified intervals.

A real-time recording tactic, made possible through the use of computer-based data collection and analysis tools, is designed to overcome these three major constraints. Using computer-based tools first allows the recording of multiple behaviors and events by specifying each behavior or event to be recorded with a numerical or letter-based key on a computer keyboard. These alphanumeric specifications may be broken down further into finer and finer discriminations through the use of various numerical notations. In addition, a qualitative narrative notation may be used to further describe atypical or unique contextual variations of a behavior or event that falls within a particular category of an observation system but is not truly representative of the typical occurrence of that behavior or event as defined. Second, computer-based tool use enables observers to record the start and stop times of all behaviors and events, providing a time-based record of when those behaviors and events actually occurred throughout an observational period. From this type of information, complex statistical/correlational formulae may be programmed to explicitly describe and analyze the many time-based associations among all behaviors and events occurring within a particular observation and related data record. Third, and perhaps most important, using computer-based tools such as that which we provide enables observers to collect their observations in real time without segmenting the total observation period and without including nonobservation time gaps in their observational tactics.

In other words, we define recording in real time as the recording of behaviors and events as they naturally occur and the recording of those behaviors and events for the entire specified observation period without taking any time out for data recording or other nonobservational tasks. Once an observer becomes familiar with the computer-based tools, to record in real time he or she is simply pressing and holding different keys on a keyboard, pressing and holding multiple keys if certain behaviors and events occur in concert, and recording multiple data files from a videotape record of the observation period to be merged and ordered later if the number of behaviors and events to be recorded and the complexity of their occurrence warrants (refer to the relevant technical guide sections for a detailed discussion of reliability procedures).

Borrowing from Roger Ray (see Ray & Delprato, 1989) and the interbehavioral descriptions by Lichtenstein (1983), real-time observational recording that is implemented according to an interbehavioral or field systems theoretical framework provides the following four steps to a more complete and inclusive description and analysis of an education or social science setting of interest. At risk of taking too much license in interpretation, we feel that Ray's steps include linguistic and topographic description, symbolic representation, and algorithmic modeling. Linguistic description involves the design and implementation of a set of terms and definitions, or category system, commonly understood by a particular group of observers and from which meaningful descriptive-analytic interpretations of particular settings stem. Topographic description involves the ways in which observations based on those linguistic descriptions are quantitatively represented, including the many measurement forms and the many ways of representing the data in graphic and tabular formats. Symbolic representation includes the ways in which we contextualize topographic descriptions through qualitative and narrative means. Finally, algorithmic modeling involves the ways in which we use statistical/correlational equations to demonstrate predictive or probabilistic relationships among the behaviors and events contained within our linguistic descriptions, and our efforts to simulate behavior-behavior and behavior-event occurrences given similar situations to what we have observed that may occur at some future point in time. In summary, through the use of a computer-based real-time approach, we can combine the appealing features of each more traditional recording tactic with all of the various traditionally recommended reporting measures *and* include sequential or time-based measures and richer contextual narratives into one preferred method.

Research and Development Directions

Use of a set of procedures and the computer-based tools to accomplish the kind of observational analysis we recommend provides quite a range of appealing research and development directions. From reading these materials and becoming familiar with the software tools that we provide, it should be clear that the observational method we recommend may enable researchers and evaluators in the education and social sciences to better understand and predict very complex behavior-behavior and behavior-event interactions. Though some contributions are becoming available, it remains in large part today as Bronfenbrenner (1979) described almost 20 years ago that education and social science literatures lack a strong database on the structure, distribution, and impact of behavior and setting event variables in context across various applied settings. In this regard, the more complex forms of behavior and the various settings in which they reside need to be more exhaustively described and understood. It also should be apparent from our argument in favor of sequential analyses that an understanding of how behaviors and events interact with one another over time may be necessary to a more inclusive description of interactive settings and may be necessary to an enhanced understanding of the meaning and nature of behavior-event relationships in situational context. More capable and more sophisticated data collection and analysis tools such as those provided by various computer applications allow us to accomplish these tasks, tasks that we could not accomplish without such tools.

In teacher education, for example, Greenwood, Carta, Arreaga-Mayer, and Rager (1991, pp. 188-189) provide a summary recommendation of the implications of the type of observational approach we provide:

> The utility of a search and validate approach to the evaluation of effective instructional practices has only just begun to be undertaken. Because of its . . . analysis of classroom behavior, and temporally related situational features of classroom instruction, it is an approach consistent with current school improvement goals. It is also an approach that focuses on the contributions classroom teachers can bring to the development of effective instructional technology. Clearly, demonstrations of the approach are warranted.

In a larger social science sense, many questions heretofore unanswered may now be investigated using a more capable observable lens. As Berliner (1992) states, this is particularly important in that a data-based, data-driven, and unabashedly quantitative approach to the analysis of fine discriminations of behavior is being provided that runs counter to the current zeitgeist that surrounds social science research in general. Though the relative merits of differing methodological approaches remain arguable, what is important is that alternatives and counterpoints to each method are provided, enabling researchers to choose among a panoply of approaches as they feel best suits a particular set of experimental questions. In this light, questions pertaining to the behavior-event differences across situations and across the developmental levels of participants may be more accurately and inclusively answered. Similarly, a tool is now available to describe and analyze the salient differences in the same participant's behavior as a function of changing situations. What is now available is a way to make these differences clear in an objective and quantifiable way.

What this method provides in addition is the flexibility to describe and document various behavior-event relationships that are unique to particular situations using different terminology that best captures a particular situation. For example, the instructional behaviors and events taking place in a public school mathematics classroom may be very different than the instructional behaviors and events that occur in a physical education classroom. Flexibility of term construction is an important feature to ensure that researchers have the full capacity to describe and analyze what they intend. Focus on a range of low-inference measures of behavior in context to high-inference measures of setting climate and atmosphere (e.g., humorous, warm, business-like, etc.) is also made simultaneously possible, and discovery of low-inference behavior changes as a function of high-inference contexts may provide valuable insights. Allowing the collection of a greater volume and variety of quantitative data, contextualizing this data in qualitative ways, and providing greater economy and flexibility of data representation should indeed provide valuable information into most interactive settings in the social sciences. Locke (1992) provides an education example that serves to illustrate:

> As I watched the individual teaching behaviors appear from beneath the moving ruler [on the sequential graphic], my attention was drawn to some unexpected juxtapositions. The teacher was modeling a skill and giving verbal instruction at the same time. This is not uncommon; teachers often recite critical verbal cues while they demonstrate—particularly when they do so in slow motion. The graphic also revealed that while she was modeling she was also observing an individual student and giving performance-specific feedback (apparently interspersed with verbal cues). Such a disparate combination of behaviors is not common at all (or so I thought), and my first response was to be incredulous. That is exactly what [this observational method] has an almost unique power to do—confront us with what is happening even when it differs sharply from our expectations. . . . I had never been conscious of this teaching move. . . . On the very next day, however, one of the videotapes used in the UMass undergraduate training program revealed several clear instances that would have matched the graphic perfectly. Had that been under my nose all the time? If so, why did it not make an impression on me? Those questions show exactly what [this observational method] can do that other methods of inquiry are far less likely to accomplish.

Clearly, much of what we discover about the world around us comes from more capable tools for observing it. Electron microscopes provide one clear example of this in that this tool helped researchers in the biological sciences to see things that they formerly could not. This phenomenon is the case in the education and social sciences as well in that more capable and inclusive observational recording instruments and observation tools that allow us to analyze behaviors and events of interest in different ways should help us see and understand behavior and event relationships that we formerly could not. The tools that we recommend provide, as Siedentop (1992) states, two appealing additions to education and social science research arsenals. First, the capability as we have stipulated before of inductively deriving category systems, with each designed for a particular setting and context, provides an appealing combination of qualitative and quantitative methodologies. Second, the ability that computer technology provides to record observations in real time eliminates traditional concerns of the relative validities of counting or timing the duration of behaviors and events. Additionally and third, as stated by Ray (1992), our tools provide alternative ways to observe behaviors and events in a quantitative way in addition to what has traditionally been available. These include, but are not limited to, regularities in behavior-event patterns (sequences of change), coherence and variability in those patterns (predictability and simplicity of sequence of change), and various velocity measures (rates of change or patterns in the rate of change).

The observational approach we recommend also infers a particular perspective on the research process. Most traditional education and social science work has a long history of deductively driven investigations based on predisposed theory. Pavlov's studies of classical conditioning provide a prime example of this approach, in which most of a researcher's efforts were spent on theory construction and the experimental documenting of the veracity of that theory. This deductive approach is in direct contrast to the observational methods we recommend. An inductive approach, in which category systems are built on the behaviors and events that actually occur in particular situations, is designed, on the other hand, to add to what we know about that situation and to add to what we know about other similar situations. In other words, the perspective we recommend is to conduct descriptive-analytic observation efforts, from those observations build an integrated knowledge base, and from that slowly building knowledge base direct further research efforts designed to uncover what we come to realize we no little about. Kantor's (1970, p. 105) pioneering theoretical framework on which we base our materials remains true today:

> Of extreme importance for the appreciation of behavior fields is their uniqueness and individuality. There is no fixed or universal type. Implied in the field construct is the principle that each class of behavior events must be analyzed according to its intrinsic [and unique] factors. . . . It is imperative to be alive to the greater complexity of non-reflex behavior, especially the interpersonal aspects of human performance.

Education and Evaluation Implications

Clearly, as education and social scientists, we are continually endeavoring to understand more about effective and not-so-effective practices in various cultural, professional, and research and training situations. The information gained from research and development activity should serve in large part to drive and to enhance our instruction and evaluation efforts. Whether new information lies in additional knowledge about new or existing behaviors and events or the new and unique ways in which we may describe and document those behaviors and events, in the ideal, these efforts should be designed to improve professional practice and to change various practices of our selected clientele in therapeutic ways. In these regards, the observational approach we have described provides great appeal in education and evaluation efforts as well.

The teacher education profession serves as an ideal illustration of the potential benefits an observational approach such as ours might provide for education and evaluation concerns. Historically, the predominant evaluation methods when observing teachers teaching in their respective environments have included tactics such as anecdotal records, narrative accounts, and Likert scale–type assessments. Inherent challenges to all of these methods have been their subjective character and their potential inclusion of user bias. In this regard, users of these more traditional assessment types have been challenged with the potential for differing evaluation outcomes within and across observers, even when observing very similar teaching situations. This has led to what Sharpe, Hawkins, and Ray (1995) refer to as an oftentimes vague and nonspecific educational process in which the important scientific components of (a) metrics of improvement discriminations over time for the same teacher-trainee and (b) a consistent metric across different teacher-trainees have been elusive.

Consider, on the other hand, the possibilities the observational techniques that we provide offer in terms of the ability to discriminate and articulate those teacher activities most influential in guiding the specified as desirable and undesirable student practices in a given subject matter or situational context. This is particularly important in relation to contemporary education reform's return to the notion of competency-based education—competencies now regularly recommended to be measured in some quantitative way. While each instructional situation observed may contain a very different set of behavior-event interactions, many situations may be described and analyzed using a similar set of terms or definitions, or the same category system. If this is the case, then the observational practices that we recommend provide a means to relatively inclusive and thoroughly quantitative and objective description of teaching practice in the context of student practice.

In addition, because teaching practices are interactive on several levels, the describing and documenting of the qualitative and interpretive character of unique or atypical coded behavior-event interactions may be included using our observational tools. Oftentimes, many more subtle but still potentially objectively described events may be narratively described given that they are atypical of a particular category system code or are more difficult to capture in coded descriptions. Again, examples of these events include such things as verbal intonations and inflections, word usage based on their connotative rather than denotative meanings, and social and physical distances that may exist between a teacher and student. What remains, however, within such a quantitative-qualitative data record to be used for evaluation purposes is that all subjective narrative descriptions are contextualized by a quantitative description of behavior and event occurrences, thereby retaining an element of objectivity and consistency to the evaluation process. Observational data such as those provided through the use of our software applications, and when coded live in the setting to be evaluated, serve as a powerfully effective immediate feedback+goal-setting tool for teachers in training (see Sharpe, 1997b; Sharpe, Hawkins, & Ray, 1995; Sharpe & Lounsbery, 1998, for a detailed procedural discussion). Such an approach also holds appeal for the education and training of a host of professionals and various clientele in settings within which behavior change is recommended.

The tools that we provide also hold great appeal for a wide variety of laboratory-type instructional experiences in which educators are interested in improving the behavior and interactions of various professionals and clientele in therapeutic ways. Setting up a prescribed set of guided observations from videotape examples using the software tools provided enables a first level of applications. Using this type of guided observation technique allows educators to help various professionals-to-be (a) learn a common observational language with which to talk over in meaningful ways the salient characteristics of effective and not-so-effective professional practices across a host of situations, (b) learn how to be accurate self- and external monitors of professional practices felt to be effective and desirable in certain situations, and (c) learn how to discriminate among effective and not-so-effective practices with a view toward the continuous improvement of their own professional practice once operating outside of the education environment. This level of applications also allows the flexibility of moving from a very simple set of only a few behaviors or events in a similar context and using only one measurement type to the gradual increasing in complexity of the number of behaviors and events, measurement types, and variable situations in which those behaviors and events are observed. The possibilities of this type of education experience are limitless, based on the collecting and archiving of a videotape library of various situations representing various types of interactions chosen for illustration.

At a second level, and pending further developments with recently available computer technologies, another variation involves computer-operated *interactive* simulation of various behavior-behavior and behavior-event interactions in context using viewer keyboard inputs to guide future video illustrations. Operating in similar fashion to children's novels that provide alternative storylines depending on reader choice functions and the turning to alternative pages, at this level of education applications, videotape viewers could first watch relatively ineffective behavior practice examples until they recognized the types of professional challenges being demonstrated. Through a CD ROM–based videotape library connected with matched behavior-event sequential data records and related conditional probabilities, viewers could then choose alternative videotape illustrations by choosing particular behaviors and events once the videotape illustration was stopped at a certain point in time. In other words, after viewing a particular situation, professionals-in-training or target clientele would be given a behavioral choice (i.e., what would you do next in this situation?), which by pressing a key on a keyboard would specify their chosen action and in turn drive the selection of the next videotape illustration to show the probable outcome of that choice given the behaviors and events occurring up to that point. Using this approach, teacher educators, for example, could teach teacher trainees relatively more effective practices in challenging situations without the potential contraindicated effects of placing them in actual classrooms before they were ready for that level of live practice. As the professional competencies of various groups being educated through this approach become relatively more important in terms of the potential negative impact on their clientele, this approach to education clearly gains in appeal. The practical education of clinical psychologists, or the pedagogical education of special educators dealing with severe and profound and violent clientele, are two of many professional training areas that could benefit from such simulated professional practice experiences, and they alleviate the potential for negative impact by professionals who may make interactive mistakes when operating in applied training experiences (see Hawkins et al., 1994; Ray, 1992; Ray & Delprato, 1989; Sharpe, Hawkins, & Lounsbery, 1998, for a more thorough discussion of simulation issues).

Though only two general areas have been discussed here related to how the observation procedures we recommend may be applied to education and evaluation concerns, providing more capable recording and analysis tools in turn provides an expanded set of possibilities when designing and implementing educational experiences. Whether the focus is on conveying information to groups of learners or providing information to be used in an evaluation capacity, we hope that the collection and relatively immediate analysis that our computer-based tools provide will facilitate a host of education and evaluation experiences that will be of benefit to various professionals-in-training and other education clientele.

Note

In closing these materials, we felt that it would be most appropriate to take from Bakeman and Gottman (1986), as we have founded a large portion of our thinking on what we feel to be seminal work and as the sequential analyses provided by our software tools are based on the mathematical algorithms that they have developed. In these regards, we find that many of us who professionally call ourselves scientists and are engaged in what we would like to think of as scientific labor find that most of the ideas and theories we use to guide our daily operations are not "new" but are more accurately characterized as "borrowed." Typically, our ideas are borrowed from significant others in our professional lives and packaged with some appeal for a previously unfamiliar audience. In addition, we borrow from what we (and others, in that the following quote has been used as a dedication for other texts devoted to behavior analysis; see Sulzer-Azaroff & Mayer, 1991) have found to be an appealing summary quote from whom most consider to be the father and pioneer of applied behavior analysis and direct observation, B. F. Skinner (1983, p. 127):

> One can picture a good life by analyzing one's feelings, but one can achieve it only by arranging environmental contingencies.

In this sense, we have attempted to provide an alternative to the subjective and qualitative forms of data gathering and analyses that currently predominate the education and social sciences. While our materials are designed to be compatible and collaborative with more qualitative methods, we will hopefully succeed in helping those who are interested in a more thoroughgoing quantitative method of describing and analyzing interactive settings of interest and in helping with the descriptive, predictive, and controlling (or arranging) stages of the scientific enterprise.

As Bakeman and Gottman (1986, pp. 200-201) so elegantly put the perspective we have attempted to bring across, and as we follow in their footsteps in our education pursuits, we would like to end with the following quote as acknowledgment:

> Observing and discovering pattern is what [we are] about. We do this kind of thing for a living, and we have chosen to do it because it is what we think science is about. Obviously we think it is not really that hard to do. But we are lonely. We want company in this enterprise. Only about 8 percent of all psychological research is based on any kind of observation. A fraction of that is programmatic research. And, a fraction of that is sequential in its thinking. This will not do. Those of us who are applying these new methods of observational research are having great success.

Endnote

1. It is important to note here that when using the software application contained within these materials to conduct reliability comparisons, a decision will need to be made regarding which formula among the three included within our software application will be used. Refer to the “Reliability Formula Summary” section following for a discussion of the relative benefits and limitations of each formula contained in our software applications.

REFERENCES

> Few truly original ideas come to pass in contemporary academe. Most of what passes for invention is merely existing ideas repackaged in appealing ways. But it is the repackaging which often moves science forward in important and insightful ways.
>
> Dennis Delprato

Altmann, S. A. (1965). Sociobiology of rhesus monkeys. II. Stochastics of social communication. *Journal of Theoretical Biology, 8*, 490-522.

Astley, C. A., Smith, O. A., Ray, R. D., Golanov, E. V., Chesney, M. A., Chalyan, V. G., Taylor, D. J., & Bowden, D. M. (1991). Integrating behavior and cardiovascular responses: The code. *American Journal of Physiology, 261*, 172-181.

Bakeman, R. (1978). Untangling streams of behavior: Sequential analyses of observation data. In G. P. Sackett (Ed.), *Observing behavior: Data collection and analysis methods* (Vol. 2, pp. 63-78). Baltimore: University Park Press.

Bakeman, R., & Gottman, J. M. (1986). *Observing interaction: An introduction to sequential analysis*. New York: Cambridge University Press.

Barlow, D. H., & Hersen, M. (1984). *Single case experimental designs: Strategies for studying behavior change* (2nd ed.). Elmsford, NY: Pergamon.

Berliner, D. C. (1992). Some perspectives on field systems research for the study of teaching expertise [Monograph]. *Journal of Teaching in Physical Education*, *12*, 96-103.

Bijou, S. W., Umbreit, J., Ghezzi, P. M., & Chao, C. (1986). Manual of instruction for identifying and analyzing referential linguistic interactions. *Psychological Record*, *36*, 491-518.

Binder, C. (1994). Measurably superior instructional methods: Do we need sales and marketing? In R. Gardner, D. M. Sainato, J. O. Cooper, T. E. Heron, W. L. Heward, J. Eshleman, & T. A. Grossi (Eds.), *Behavior analysis in education: Focus on measurably superior instruction* (pp. 21-31). Pacific Grove, CA: Brooks/Cole.

Bronfenbrenner, U. (1979). Contexts of child rearing: Problems and prospects. *American Psychologist*, *34*, 84-89.

Brown, S. R. (1980). *Political subjectivity: Applications of Q methodology in political science*. London: Yale University Press.

Buchler, J. (Ed.). (1955). *Philosophical writings of Peirce*. New York: Dover.

Carnine, D. W., & Fink, W. T. (1978). Increasing the rate of presentation and use of signals in elementary classroom teachers. *Journal of Applied Behavior Analysis*, *11*, 35-46.

Chatfield, C., & Lemon, R. E. (1970). Analyzing sequences of behavioural events. *Journal of Theoretical Biology*, *29*, 427-445.

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

Cohen, J. (1965). Some statistical issues in psychological research. In B. B. Wolman (Ed.), *Handbook of clinical psychology.* New York: McGraw-Hill.

Cooper, J. O., Heron, T. E., & Heward, W. L. (1987). *Applied behavior analysis*. Toronto, Canada: Merrill.

Cooper, M. L., Thomson, C. L., & Baer, D. M. (1970). The experimental modification of teacher attending behavior. *Journal of Applied Behavior Analysis*, *3*, 153-157.

Cossairt, A., Hall, R. V., & Hopkins, B. L. (1973). The effects of experimenter's instructions, feedback, and praise on teacher praise and student attending behavior. *Journal of Applied Behavior Analysis*, *6*, 89-100.

Croll, P. (1986). *Systematic classroom observation*. Philadelphia: Falmer.

Darst, P. W., Zakrajsek, D. B., & Mancini, V. H. (Eds.). (1989). *Analyzing physical education and sport instruction*. Champaign, IL: Human Kinetics.

Davison, M., & McCarthy, D. (1988). *The matching law: A research review.* Hillsdale, NJ: Lawrence Erlbaum.

Delprato, D. J. (1992). Behavior field systems analysis: History and scientific relatives [Monograph]. *Journal of Teaching in Physical Education*, *12*, 3-8.

Doyle, W. (1990). Themes in teacher education research. In W. R. Houston, M. Haberman, & J. Sikula (Eds.), *Handbook of research on teacher education* (pp. 3-24). New York: Macmillan.

Dunkin, M. J., & Biddle, B. J. (1974). *The study of teaching*. New York: Holt, Rinehart & Winston.

Dwyer, D. (1996). We're in this together. *Educational Leadership*, *54*(3), 24-26.

Einstein, A., & Infeld, L. (1938). *The evolution of physics*. New York: Simon & Schuster.

Ekman, P. W., & Friesen, W. (1978). *Manual for the facial action coding system*. Palo Alto, CA: Consulting Psychologist.

Ellul, J. (1964). *The technological society.* New York: Knopf.

Erickson, F. (1982). The analysis of audiovisual records as a primary data source. In A. Grimshaw (Ed.), Sound-image records in social interaction research [Special Issue]. *Journal of Sociological methods and Research*, *11*(12), 213-232.

Faraone, S. V. (1983). The behavior as language analogy: A critical examination and application of conversational interaction. *Behaviorism*, *11*, 27-43.

Flanders, N. A. (1970). *Analyzing teacher behavior.* Reading, MA: Addison-Wesley.

Friman, P. C., Wilson, K. G., & Hayes, S. C. (1998). Behavior analysis of private events is possible, progressive, and nondualistic: A response to Lamal. *Journal of Applied Behavior Analysis*, *31*, 707-708.

Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children*, *53*, 199-208.

Glaser, B., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.

Gottman, J. M. (1979a). Detecting cyclicity in social interactions. *Psychological Bulletin*, *86*, 338-348.

Gottman, J. M. (1979b). *Marital interaction: Experimental investigations*. New York: Academic Press.

Gottman, J. M., & Roy, A. K. (1990). *Sequential analysis: A guide for behavioral researchers*. New York: Cambridge University Press.

Greenwood, C. R., Carta, J. J., Arreaga-Mayer, C., & Rager, A. (1991). The behavior analyst consulting model: Identifying and validating naturally effective instructional methods. *Journal of Behavioral Education, 1,* 165-191.

Greenwood, C. R., Carta, J. J., & Atwater, J. (1991). Ecobehavioral analysis in the classroom: Review and implications. *Journal of Behavioral Education*, *1*, 59-77.

Greenwood, C. R., Delquadri, J. C., Stanley, S. O., Terry, B., & Hall, R. V. (1985). Assessment of eco-behavioral interaction in school settings. *Behavioral Assessment*, *7*, 331-347.

Greer, R. D. (1985). *Handbook for professional change agents at the Margaret Chapman School*. Hawthorne, NY: The Margaret Chapman School.

Gresham, F. M., Gansle, K. A., & Noell, G. H. (1993). Treatment integrity in applied behavior analysis with children. *Journal of Applied Behavior Analysis*, *26*, 257-264.

Hall, R. V., Panyon, M., Rabon, D., & Broden, M. (1968). Instructing beginning teachers in reinforcement procedures which improve classroom control. *Journal of Applied Behavior Analysis, 1,* 315-322.

Hawkins, A., & Sharpe, T. L. (Eds.). (1992). Field systems analysis: An alternative for the study of teaching expertise. *Journal of Teaching in Physical Education*, *12*, 1-131.

Hawkins, A., Sharpe, T. L., & Ray, R. (1994). Toward instructional process measurability: An interbehavioral field systems perspective. In R. Gardner, D. M. Sainato, J. O. Cooper, T. E. Heron, W. L. Heward, J. Eshleman, & T. A. Grossi (Eds.), *Behavior analysis in education: Focus on measurably superior instruction* (pp. 241-255). Pacific Grove, CA: Brooks/Cole.

Hawkins, A. H., Wiegand, R. L., & Landin, D. K. (1985). Cataloguing the collective wisdom of teacher educators. *Journal of Teaching in Physical Education*, *4*, 241-255.

Hawkins, R. P. (1982). Developing a behavior code. In D. P. Hartmann (Ed.), *Using observers to study behavior: New directions for methodology of social and behavior science* (pp. 21-35). San Francisco: Jossey-Bass.

Hawkins, R. P., & Dobes, R. W. (1977). Behavioral definitions in applied behavior analysis: Explicit or implicit. In B. C. Etzel, J. M. LeBlanc, & D. M. Baer (Eds.), *New developments in behavioral research: Theory, methods, and applications. In honor of Sidney W. Bijou*. Hillsdale, NJ: Lawrence Erlbaum.

Heward, W. L., & Cooper, J. O. (1992). Radical behaviorism: A productive and needed philosophy for education. *Journal of Behavioral Education*, *24*, 345-365.

Hunt, P., Alwell, M., Farron-Davis, F., & Goetz, L. (1996). Creating socially supportive environments for fully included students who experience multiple disabilities. *Journal of The Association for Persons With Severe Handicaps*, *21*, 53-71.

Ingham, P., & Greer, R. D. (1992). Changes in student and teacher responses in observed and generalized settings as a function of supervisor observations. *Journal of Applied Behavior Analysis*, *25*, 153-164.

Issues in Interobserver Reliability. (1977). *Journal of Applied Behavior Analysis*, *10*(2).

Jacobson, N. S., & Anderson, E. A. (1982). Interpersonal skill and depression in college students: An analysis of the timing of self-disclosures. *Behavior Therapy*, *13*, 271-282.

Johnson, H., Blackhurst, A. E., Maley, K., Bomba, C., Cox-Cruey, T., & Dell, A. (1995). Development of a computer-based system for the unobtrusive collection of direct observational data. *Journal of Special Education Technology*, *12*, 291-300.

Johnson, S. M., & Bolstad, O. D. (1973). Methodological issues in naturalistic observation: Some problems and solutions for field research. In L. A. Hamerlynck, L. C. Handy, & E. J. Mash (Eds.), *Behavior change: Methodology, concepts, and practice* (pp. 7-67). Champaign, IL: Research Press.

Johnston, J. M., & Pennypacker, H. S. (1980). *Strategies and tactics of human behavioral research*. Hillsdale, NJ: Lawrence Erlbaum.

Kamps, D. M., Leonard, B. R., Dugan, E. P., Boland, B., & Greenwood, C. R. (1991). The use of ecobehavioral assessment to identify naturally occurring effective procedures in classrooms serving students with autism and other developmental disabilities. *Journal of Behavioral Education*, *1*, 367-397.

Kantor, J. R. (1953). *The logic of modern science*. Chicago: Principia.

Kantor, J. R. (1959). *Interbehavioral psychology*. Granville, OH: Principia.

Kantor, J. R. (1969). *The scientific evolution of psychology* (Vol. 2). Chicago: Principia.

Kantor, J. R. (1970). An analysis of the experimental analysis of behavior (TEAB). *Journal of the Experimental Analysis of Behavior, 13*, 101-105.

Kantor, J. R. (1977). *Psychological linguistics*. Chicago: Principia.

Kazdin, A. E. (1982). *Single case research designs*. New York: Oxford University Press.

Kerlinger, F. N. (1986). *Foundations of behavioral research* (3rd ed.). New York: Holt, Rinehart & Winston.

Landin, D. K., Hawkins, A. H., & Wiegand, R. L. (1986). Validating the collective wisdom of teacher educators. *Journal of Teaching in Physical Education*, *5*, 252-271.

LeCompte, M. D., & Preissle, J. (1993). *Ethnography and qualitative design in educational research* (2nd ed.). San Diego, CA: Academic Press.

Lichtenstein, P. E. (1983). The interbehavioral approach to psychological theory. In N. W. Smith, P. T. Mountjoy, & D. H. Ruben (Eds.), *Reassessment in psychology: The interbehavioral alternative* (pp. 3-20). Washington, DC: University Press of America.

Light, J., Collier, B., & Parnes, P. (1985a). Communicative interaction between young nonspeaking physically disabled children and their primary caregivers: Part I—Discourse patterns. *Augmentative and Alternative Communication*, *1*, 74-83.

Light, J., Collier, B., & Parnes, P. (1985b). Communicative interaction between young nonspeaking physically disabled children and their primary caregivers: Part II—Communicative functions. *Augmentative and Alternative Communication*, *1*, 98-107.

Light, J., Collier, B., & Parnes, P. (1985c). Communicative interaction between young nonspeaking physically disabled children and their primary caregivers: Part III—Modes of communication. *Augmentative and Alternative Communication*, *1*, 125-133.

Locke, L. F. (1992). Field systems research: Sport pedagogy perspectives [Monograph]. *Journal of Teaching in Physical Education*, *12*, 85-89.

Martens, B. K., & Witt, J. C. (1988a). Ecological behavioral analysis. In M. Hersen, R. M. Eisler, & P. M. Miller (Eds.), *Progress in behavior modification* (Vol. 27, pp. 115-140). Newbury Park, CA: Sage.

Martens, B. K., & Witt, J. C. (1988b). Expanding the scope of behavioral consultation: A systems approach to classroom change. *Professional School Psychology*, *3*, 271-281.

McSweeney, F. K., Farmer, V. A., Dougan, J. D., & Whipple, J. E. (1986). The generalized matching law as a description of multiple-schedule responding. *Journal of the Experimental Analysis of Behavior*, *45*, 83-101.

Metzler, M. (1989). A review of research on time in sport pedagogy. *Journal of Teaching in Physical Education*, *8*, 87-103.

Miles, M. B., & Huberman, A. M. (1984). *Qualitative data analysis: A sourcebook of new methods.* Beverly Hills, CA: Sage.

Miller, S. P., Harris, C., & Watanabe, A. (1991). Professional coaching: A method for increasing effective and decreasing ineffective teacher behaviours. *Teacher Education and Special Education*, *14*, 183-191.

Mjrberg, A. A. (1972). Ethology of the bicolor damselfish, *Eupomaclatsus partitus (Pisces Pomacentridae)*: A comparative analysis of laboratory and field behaviour. *Animal Behavior Monographs*, *5*.

Morris, E. K. (1984). Public information, dissemination, and behavior analysis. *The Behavior Analyst*, *8*, 95-110.

Morris, E. K. (1991). Deconstructing "technological to a fault." *Journal of Applied Behavior Analysis*, *24*, 411-416.

Morris, E. K. (1992). The aim, progress, and evolution of behavior analysis. *The Behavior Analyst*, *15*, 3-29.

Myerson, J., & Hale, S. (1988). Choice in transition: A comparison of melioration and the kinetic model. *Journal of the Experimental Analysis of Behavior*, *49*, 291-302.

Neitzche, F. W. (1978). *Thus spake Zarathustra*. New York: Penguin Books.

Newman, B. (1992). *The reluctant alliance: Behaviorism and humanism*. Buffalo, NY: Prometheus.

Odom, S. L., & Haring, T. G. (1994). Contextualism and applied behavior analysis: Implications for early childhood education for children with disabilities. In R. Gardner, D. M. Sainato, J. O. Cooper, T. E. Heron, W. L. Heward, J. Eshleman, & T. A. Grossi (Eds.), *Behavior analysis in education: Focus on measurably superior instruction* (pp. 87-99). Pacific Grove, CA: Brooks/Cole.

O'Reilly, M. F., & Renzaglia, A. (1994). A systematic approach to curriculum selection and supervision strategies: A preservice practicum supervision model. *Teacher Education and Special Education*, *17*, 170-180.

Page, T. J., Iwata, B. A., & Reid, D. H. (1982). Pyramidal training: A large scale application with institutional staff. *Journal of Applied Behavior Analysis*, *15*, 355-352.

Penman, R. (1980). *Communication processes and relationships*. London: Academic Press.

Peterson, L., Homer, A., & Wonderlich, S. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis*, *15*, 477-492.

Pett, M. A., Vaughan-Cole, B., Egger, M., & Dorsey, P. (1988). Wrestling meaning from interactional data: An empirically-based strategy for deriving multiple molar constructs in parent-child interaction. *Behavioral Assessment*, *10*, 299-318.

Ray, R. D. (1983). Interbehavioral systems, temporal settings and organismic health. In N. W. Smith, P. T. Mountjoy, & D. H. Ruben (Eds.), *Reassessment in psychology: The interbehavioral alternative* (pp. 361-380). Washington, DC: University Press of America.

Ray, R. D. (1992). Interbehavioral methodology: Lessons from simulation [Monograph]. *Journal of Teaching in Physical Education*, *12*, 105-114.

Ray, R. D., & Delprato, D. J. (1989). Behavioral systems analysis: Methodological strategies and tactics. *Behavioral Science*, *34*, 81-127.

Ray, R. D., Upson, J. D., & Henderson, B. J. (1977). A systems approach to behavior III: Organismic pace and complexity in time-space fields. *Psychological Record*, *27*, 649-682.

Rechsly, D. J., & Wilson, M. S. (1996). Assessment in school psychology training and practice. *School Psychology Review*, *25*, 9-23.

Rodger, R. S., & Rosebrugh, R. D. (1979). Computing a grammar for sequences of behavioral acts. *Animal Behavior*, *27*, 737-749.

Rosenshine, B. V., & Furst, N. (1973). The use of direct observation to study teaching. In R.M.W. Travers (Ed.), *Second handbook of research on teaching* (pp. 122-183). Chicago: Rand McNally.

Ruben, D. H., & Delprato, D. J. (Eds.). (1987). *New ideas in therapy.* Westport, CT: Greenwood.

Sackett, G. P. (1979). The lag sequential analysis of contingency and cyclicity in behavioral interaction research. In J. D. Osofsky (Ed.), *Handbook of infant development* (pp. 623-649). New York: John Wiley.

Sackett, G. P. (1980). Lag sequential analysis as a data reduction technique in social interaction research. In D. B. Sawin, R. C. Hawkins II, L. O. Walker, & J. H. Penticuff (Eds.), *Exceptional infant* (Vol. 4, pp. 300-340). New York: Brunner-Mazel.

Scheflen, A. E. (1982). Comments on the significance of interaction rhythms. In M. Davis (Ed.), *Interaction rhythms* (pp. 13-22). New York: Human Sciences Press.

Sharpe, T. L. (1997a). An introduction to sequential behavior analysis and what it offers physical education teacher education researchers. *Journal of Teaching in Physical Education*, *16*, 368-375.

Sharpe, T. L. (1997b). Using technology in preservice teacher supervision. *The Physical Educator*, *54*, 11-19.

Sharpe, T. L., Brown, M., & Crider, K. (1995). The effects of a sportsmanship curriculum intervention on generalized positive social behavior of urban elementary school students [Special Section]. *Journal of Applied Behavior Analysis*, *28*, 401-416.

Sharpe, T. L., Crider, K., Vyhlidal, T., & Brown, M. (1996). Description and effects of prosocial instruction in an elementary physical education setting. *Education and Treatment of Children*, *19*, 435-457.

Sharpe, T. L., Harper, W., & Brown, S. (1998). In response: Further reflections on technology, science, and culture. *Quest*, *50*, 332-343.

Sharpe, T. L., & Hawkins, A. (1992). Expert and novice elementary specialists: A comparative analysis [Monograph]. *Journal of Teaching in Physical Education*, *12*, 55-75.
Sharpe, T. L., & Hawkins, A. (1998). Technology and the information age: A cautionary tale for higher education. *Quest*, *50*, 19-32.
Sharpe, T. L., Hawkins, A., & Lounsbery, M. (1998). Using technology to study and evaluate human interaction: Practice and implications of a sequential behavior approach. *Quest*, *50*, 389-401.
Sharpe, T. L., Hawkins, A., & Ray, R. (1995). Interbehavioral field systems assessment: Examining its utility in preservice teacher education. *Journal of Behavioral Education*, *5*, 259-280.
Sharpe, T. L., & Lounsbery, M. (1998). The effects of a sequential behavior analysis protocol on the teaching practices of undergraduate trainees. *School Psychology Quarterly*, *12*, 327-343.
Sharpe, T. L., Lounsbery, M., & Bahls, V. (1997). Description and effects of sequential behavior practice in teacher education. *Research Quarterly for Exercise and Sport*, *68*, 222-232.
Sharpe, T. L., Spies, R., Newman, R., & Spickelmier-Vallin, D. (1996). Assessing and improving the accuracy of inservice teachers perceptions of daily practice. *Journal of Teaching in Physical Education*, *15*, 297-318.
Shriver, M. D., & Kramer, J. J. (1997). Application of the generalized matching law for description of student behavior in the classroom. *Journal of Behavioral Education*, *7*, 131-149.
Siedentop, D. (1992). New folks in the neighborhood: A sport pedagogy perspective [Monograph]. *Journal of Teaching in Physical Education*, *12*, 90-95.
Silverman, S., & Solmon, M. (1998). The unit of analysis in field research: Issues and approaches to design and data analysis. *Journal of Teaching in Physical Education*, *17*, 270-284.
Skinner, B. F. (1968). *The technology of teaching*. New York: Appleton-Century-Crofts.
Skinner, B. F. (1983). *Notebooks*. Englewood Cliffs, NJ: Prentice Hall.
Skinner, B. F. (1989). *Recent issues in the analysis of behavior.* Toronto, Canada: Merrill.
Sprague, J. R., & Horner, R. H. (1992). Covariation within functional response classes: Implications for treatment of severe problem behavior. *Journal of Applied Behavior Analysis*, *25*, 735-745.
Sprague, J. R., & Horner, R. H. (1994). Covariation within functional response classes: Implications for treatment of severe problem behavior. In T. Thompson & D. B. Gray (Eds.), *Destructive behavior in developmental disabilities: Diagnosis and treatment*. Sage focus editions, Vol. 170 (pp. 213-242). Thousand Oaks, CA: Sage.
Stallings, J., Needels, M., & Sparks, G. M. (1987). Observation for the improvement of classroom learning. In D. Berliner & B. Rosenshine (Eds.) *Talks to teachers* (pp. 129-158). New York: Random House.
Sulzer-Azaroff, B., & Mayer, G. R. (1991). *Behavior analysis for lasting change*. New York: Harcourt Brace Jovanovich College Publishers.
Touchette, P. E., MacDonald, R. F., & Langer, S. N. (1985). A scatter plot for identifying stimulus control of problem behavior. *Journal of Applied Behavior Analysis*, *18*, 343-351.
Utley, B. L., Zigmond, N., & Strain, P. S. (1987). How various forms of data affect teacher analysis of student performance. *Exceptional Children*, *53*, 411-422.

van der Mars, H. (1989). Systematic observation: An introduction. In P. W. Darst, D. B. Zakrajsek, & V. H. Mancini (Eds.), *Analyzing physical education and sport instruction* (pp. 3-17). Champaign, IL: Human Kinetics.

Wahler, R. G., & Hann, D. H. (1987). An interbehavioral approach to clinical child psychology: Toward understanding troubled families. In D. H. Ruben & D. J. Delprato (Eds.), *New ideas in therapy* (pp. 53-78). Westport, CT: Greenwood.

Wampold, B. E. (1986). State of the art in sequential analysis: Comment on Lichtenberg and Heck. *Journal of Counseling Psychology*, *33*, 182-185.

Wampold, B. E. (1992). The intensive examination of social interactions. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis: New directions for psychology and education* (pp. 93-131). Hillsdale, NJ: Lawrence Erlbaum.

Warger, C. L., & Aldinger, L. E. (1984). Improving teacher supervision: The preservice consultation model. *Teacher Education and Special Education*, *7*, 155-163.

Watson, J. B. (1970). *Behaviorism*. New York: Norton.

Willems, E., & Raush, H. (Eds.). (1980). *Naturalistic viewpoints on psychological research.* New York: Holt, Rinehart & Winston.

Witt, J. C., Noell, G. H., Lafleur, L. H., & Mortenson, B. P. (1997). Teacher use of interventions in general education settings: Measurement and analysis of the independent variable. *Journal of Applied Behavior Analysis*, *30*, 693-696.