

## BAT 3.3

[Behavior Analyst Today](#)  
Volume 3 Issue 3

## THE BEHAVIOR ANALYST TODAY

## A Context for Science with a Commitment to Behavior Change

## TABLE OF CONTENTS

<b>Greenway &amp; Wulfert:</b> The Current Status and Future Direction of Clinical Behavior Analysis: Introduction to the Special Section	246	<b>Butler, Weaver, Doggett &amp; Watson:</b> Countering Teacher Resistance in Behavioral Consultation: Recommendations for the School-Based Consultant	282
<b>Kohlenberg, Bolling, Kanter &amp; Parker:</b> Clinical Behavior Analysis: Where It Went Wrong, How It Was Made Good Again, and Why Its Future is So Bright.	248	<b>LeBlanc, Hagopian, Maglieri &amp; Poling:</b> Decreasing the Intensity of Reinforcement-Based Interventions for Reducing Behavior: Conceptual Issues and a Proposed Model for Clinical Practice	289
<b>Wulfert:</b> Can Contextual Therapies Save Clinical Behavior Analysis?	254	<b>Hyten:</b> On the Identity Crisis in OBM	301
<b>Gifford:</b> Socrates and the Dodo Bird: Clinical Behavior Analysis and Psychotherapy Research	259	<b>Gilbreath &amp; Harris:</b> Performance-Based Pay in the Workplace: Magic Potion or Malevolent Poison?	311
<b>Williams:</b> Constructing a Behavior Analytical Helping Process	262	<b>Bryington, Palmer &amp; Watson:</b> The Estimation of Interobserver Agreement in Behavioral Assessment	323
<b>Augustson:</b> An Overview of Some Current Challenges within the Field of Clinical Behavior Analysis	265	<b>Collins:</b> Using Peers to Facilitate Learning by Students with Moderate Disabilities	329
<b>Skinner, Neddenriep, Bradley-Klug &amp; Ziemann:</b> Advances in Curriculum-Based Measurement: Alternative Rate Measures for Assessing Reading Skills in Pre- and Advanced Readers	270	<b>Keenen &amp; Kerr:</b> Human time-based schedule performance: Exploring interactions between instructions and response-reinforcer contiguity	342

**THE BEHAVIOR ANALYST TODAY:  
PUBLISHER'S STATEMENT  
VOLUME 3, ISSUE 3**

**Published October 22, 2002**

The Behavior Analyst Today (BAT) is published quarterly by Joseph Cautilli. BAT is an online, electronic publication of general circulation to the scientific community. BAT's mission is to provide a concentrated behavior analytic voice among voices which are more cognitive and structural. BAT emphasizes functionalism and behavioral approaches to verbal behavior. Additionally, BAT hopes to highlight the importance of conducting research from a strong theoretical base. BAT areas of interest include, but are not limited to Clinical Behavior Analysis, Behavior Models of Child Development, Community based behavioral analytic interventions, and Behavioral Philosophy. BAT is an independent publication and is in no way affiliated with any other publications.

The materials, articles, and information provided on this website have been prepared by the staff of the Behavior Analyst Today for informational purposes only. The information contained in this web site is not intended to create any kind of patient-therapist relationship or representation whatsoever.

For a free subscription to The Behavior Analyst Today, send the webmaster an e-mail containing your name, e-mail address, and the word "subscribe" in the subject box, and you will be added to the subscription list. You will receive notice of publication of each new issue via e-mail that will contain a hyperlink to the latest edition. You may also subscribe to the BAT journal by visiting <http://www.behavior-analyst-online.org>.

**Lead Co-Editors:**

[Joseph Cautilli, M.Ed., BCBA](#)  
[Beth Rosenwasser, M.Ed., BCBA, CAC](#)

**Senior Associate Editor:**

[Margaret Hancock, M.Ed., BCBA](#)

**Associate Editors:**

[Michael Weinberg, Ph.D., BCBA](#)  
[John Eshleman, Ed.D.](#)  
[David R. Feeney, Ed.D.](#)

BAT welcomes Dr. David Feeney in two new roles: Associate Editor and Chief Layout Editor. Our special thanks to Dr. Feeney for his expertise in both technology and behavior analysis, allowing him to assume both positions simultaneously. Contact David Feeney at [DavidFeeney@aol.com](mailto:DavidFeeney@aol.com) or <http://foxonline.temple.edu/davidfeeney>

**EDITORIAL BOARD**

Jack Apsche, Ed.D., ABPP  
 Thomas Critchfield, Ph.D.  
 Michael Dougher, Ph.D.  
 David Greenway, Ph.D.  
 Raymond Reed Hardy, Ph.D.  
 Lee Kern, Ph.D.  
 Mareile Koenig, Ph.D.  
 Richard Kubina, Ph.D.  
 Stephen Ledoux, Ph.D.  
 Ethan Long, Ph.D.  
 Stein Lund  
 Frances McSweeney, Ph.D.  
 Sherry Milchick, M.Ed., BCBA  
 Edward K. Morris, Ph.D.  
 Daniel J. Moran, Ph.D.  
 John T. Neisworth, Ph.D.  
 Martha Pelaez, Ph.D.  
 Lillian Pelios, Ph.D.  
 Patrick Progar, Ph.D.  
 David Reitman, Ph.D.  
 Chris Riley-Tillman, Ph.D.  
 Lynn Santilli Connor, MSW, LSW, BCBA  
 Sherry Serdikoff, Ph.D., BCBA  
 Janet Sloand Armstrong, M.Ed., BCBA  
 Ralph Spiga, Ph.D.  
 Richard Weissman, Ph.D., BCBA

**Submission Information:**

Most contributions are by invitation and all are then peer-reviewed and edited. The editors, however, welcome unsolicited manuscripts, in which case, we suggest potential authors send an abstract or short summary of contents and we will respond as to our interest in a full manuscript submission. In all cases, manuscripts should be submitted electronically saved in "rich text format" (.rtf) to BOTH Beth Rosenwasser at [ibrosie@aol.com](mailto:ibrosie@aol.com) and Joe Cautilli at [jcautill@astro.temple.edu](mailto:jcautill@astro.temple.edu). Please adhere to APA format and use "Times New Roman" font in 11 pt. Throughout. In references, however, please italicize the places where APA format would have you underline. Headings are encouraged and must follow APA format.

**Our Mission**

The Behavior Analyst Today is committed to increasing the communication between the sub disciplines within behavior analysis, such as behavioral assessment, work with various populations, basic and applied research. Through achieving this goal, we hope to see less fractionation and greater cohesion within the field. The Behavior Analyst Today strives to be a high quality journal, which also brings up to the minute information on current developments within the field to those who can benefit from those developments. Founded as a newsletter for master level practitioners in Pennsylvania and those represented in the clinical behavior analysis SIG at ABA and those who comprised the BA SIG at the Association for the Advancement of Behavior Therapy, BAT has evolved to being a primary form of communication between researchers and practitioners, as well as a primary form of communication for those outside behavior analysis. Thus the Behavior Analyst Today will continue to publish original research, reviews of sub disciplines, theoretical and conceptual work, applied research, program descriptions, research in organizations and the community, clinical work, and curriculum developments. In short, we strive to publish all which is behavior analytic. Our vision is to become the voice of the behavioral community.

## THE CURRENT STATUS AND FUTURE DIRECTION OF CLINICAL BEHAVIOR ANALYSIS: INTRODUCTION TO THE SPECIAL SECTION

David E. Greenway  
University of Louisiana at Lafayette  
Edelgard Wulfert  
University at Albany, State University of New York

During the 1960s and early 1970s, when Skinnerian behaviorism had reached its apogee, many behavior therapists in clinical practice employed operant principles in their work with the range of clinical disorders. Behavior therapists prided themselves in the utilization of principles that were derived from well-controlled laboratory experiments. Their empirically based approach commanded a far greater scientific respectability than the rather vaguely defined interventions of psychodynamic and humanistic therapists. However, starting in the early 1970s, cognitive psychology began to develop momentum and many behavior therapists gradually began to shift from a behavioral to a cognitive stance. This movement eventually resulted in a split between two types of behaviorally oriented practitioners. On the one hand, there were the applied behavior analysts who faithfully carried on the operant tradition. They worked mainly with developmentally disabled children and adults or employed behavior modification principles in relatively closed systems such as educational, correctional, and organizational settings. On the other hand, there were the behavior therapists who worked with the range of problems presented by adult outpatients. In their majority, these therapists began to embrace the new cognitive-behavioral perspective, whereas psychotherapists employing Skinnerian principles in the treatment of adult outpatients all but disappeared.

This situation changed when in the late 1980s two approaches to adult outpatient psychotherapy were introduced which were both based on Skinner's radical behaviorism: Acceptance and Commitment Therapy (ACT; Hayes, 1987) and Functional Analytic Psychotherapy (FAP; Kohlenberg & Tsai, 1987). With the advent of these therapies, a novel form of operant based behavior therapy was born, which during the early 1990s eventually established itself as 'clinical behavior analysis'

(CBA). CBA has its philosophical roots in a modernized version of radical behaviorism termed 'functional contextualism' (Biglan & Hayes, 1996) and is geared toward the amelioration of emotional problems that more typically have been the purview of traditional adult psychotherapy.

A number of writers have shown how CBA based approaches can be extended to the treatment of a wide spectrum of psychological disorders, including problems as complex as personality, mood, sexual, or substance use disorders (Dougher, 1999; Hayes, Jacobson, Follette, & Dougher, 1994). The beauty of CBA is that it deals effectively with the fundamental theoretical problem of how talking within the therapy session helps ameliorate clients' problems that occur in their everyday lives. By carefully examining the verbal interactions between therapists and clients within a conceptual framework that brings to bear the principles of verbal behavior (Skinner, 1957), rule-governed behavior (Skinner, 1969), and stimulus equivalence (Sidman, 1994), clinical behavior analysts have developed a theoretical basis for adult psychotherapy that is consistent with contemporary radical behaviorism. Thus, for the first time in a long time there is evidence of a renewed rapprochement between basic behavioral science and its application. Over the past several decades, behavior therapy has gradually separated itself from its theoretical and philosophical foundation in behaviorism, which during its beginnings had been its greatest strength. With the advent of CBA, the gap between the basic and the applied branch of the behavioral approach is closing. Thus, CBA holds the clear promise of reestablishing the conceptual integrity of the behavioral approach to psychotherapy.

Given this promise, it is understandable that CBA has generated a great deal of enthusiasm and excitement among behavior

therapists. It is clear that CBA enjoys a rising popularity when one examines the annual convention programs of behavior therapists' most important professional organization, the Association for Advancement of Behavior Therapy (AABT). Over the past several years, ACT and FAP related topics have been increasingly referenced in AABT convention programs; indeed, 'acceptance' has even garnered a separate subject heading.

The excitement over CBA as a therapeutic approach with a solid conceptual foundation may in part reflect the disillusionment of some cognitive behavioral therapists with their own approach, given that recent controlled studies have failed to establish its superiority over non-behavioral therapies. As CBA views the verbal transactions between client and therapist in contextualized operant terms, it focuses our attention on how behavior *therapists* rather than *therapies* instigate change in their clients' behavior. By focusing attention on the effects of verbal behavior, clinical behavior analysts seem well equipped to generate a more parsimonious and useful way of conceptualizing how therapy works.

The idea for the papers presented in this special section grew out of a panel discussion that one of us (D.G.) arranged under the title 'Current Status and Future Direction of Clinical Behavior Analysis' at the 24<sup>th</sup> annual meeting of the Association for Behavior Analysis (ABA) in 1998. An informal presentation of the ideas that now have been formalized led to an animated dialogue with ABA members. The following

papers were written in a spirit of intellectual honesty, highlighting both the strengths and weaknesses of CBA. It is our hope that the papers will not only stimulate debate among readers but also provide guidance for needed empirical work, which we consider vital to the future development and health of CBA.

## REFERENCES

- Biglan, A., & Hayes, S. C. (1996). Should the behavioral sciences become more pragmatic? The case for functional contextualism in research on human behavior. *Applied and Preventive Psychology: Current Scientific Perspectives*, 5, 47-57.
- Dougher, M. J. (Ed.) (1999). *Clinical behavior analysis*. Reno, NV: Context Press.
- Hayes, S. C. (1987). A contextual approach to therapeutic change. In N. S. Jacobson (Ed.), *Psychotherapists in clinical practice*. New York: Guilford.
- Hayes, S. C., Jacobson, V. M., Follette, V. M., & Dougher, M. J. (Eds.) (1994). *Acceptance and change: Content and context in psychotherapy* (pp. 203-217). Reno, NV: Context Press.
- Kohlenberg, R. J., & Tsai, M. (1987). Functional analytic psychotherapy. In N. S. Jacobson (Ed.), *Psychotherapists in clinical practice*. New York: Guilford.
- Sidman, M. (1994). *Equivalence relations and behavior: A research story*. Boston: Author's Cooperative.
- Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.
- Skinner, B. F. (1969). *Contingencies of reinforcement: A theoretical analysis*. New York: Appleton-Century-Crofts.

**Author note:** Address editorial correspondence to: David Greenway, Ph.D., P.O. Box 43131, Psychology Department, University of Louisiana at Lafayette, Lafayette, LA 70504-3131. Email: [greenway@louisiana.edu](mailto:greenway@louisiana.edu)

## CLINICAL BEHAVIOR ANALYSIS: WHERE IT WENT WRONG, HOW IT WAS MADE GOOD AGAIN, AND WHY ITS FUTURE IS SO BRIGHT.

Robert J. Kohlenberg, Madelon Y. Bolling, Jonathan W. Kanter & Chauncey R. Parker  
University of Washington

This paper traces the birth, quiescence and renaissance of clinical behavior analysis (CBA). CBA is the application of radical behaviorism to outpatient adult behavior therapy. It addresses the question of how talking in the consulting room helps the client outside of the office, in his or her daily life. The answer as formulated by CBA has led to exciting and significant developments with considerable promise for improving therapeutic interventions. A brief historical account of CBA is described that involves the interplay of three strands involving clinical applications of behaviorism: behavior therapy, applied behavior analysis, and the development of the Association for the Advancement of Behavior Therapy (AABT). These strands are traced through publications in *Behavior Therapy* from its inception to the present. We contend that there is a need in AABT and in behavior therapy in general for what CBA has to offer. As we see it, the major problems facing the AABT membership with its current emphasis on cognitive therapy and empirically validated treatments include the lack of a coherent theoretical base that can embrace all of the techniques used by behavior therapists. Now with all the behavioral procedures that have been developed, a horrendous question arises, "When do you use which procedure for what kind of person?" We conclude that far from being a thing of the past, CBA has a bright future in answering this question. Behavior analysis of the therapeutic situation offers a unique, coherent theoretical base that can embrace all techniques used by behavior therapists, including cognitive therapy strategies.

---

Clinical behavior analysis (CBA) is defined as the application of radical behaviorism (Skinner 1953, 1974) to answer the most basic question about outpatient adult behavior therapy (or any other type of psychotherapy) (Kohlenberg, Tsai & Dougher, 1993). Since outpatient treatment consists of verbal interchanges<sup>1</sup> between client and therapist, the question is this: what is the mechanism that explains how this talking helps the client outside of the office in his or her daily life? In this paper, we contend that CBA is an exciting, new, and significant development that holds considerable promise for improving therapeutic interventions. We also recognize that most behavior therapists are only superficially familiar, if at all, with CBA and are not aware of its considerable potential as a highly effective treatment. There are several factors that account for the relative invisibility of CBA, not least of which is its mercurial appearance over the last 46 years.

### THE BIRTH OF CBA

In *Science and Human Behavior* (1953), Skinner gave an analysis of psychotherapy,

---

<sup>1</sup> As we define it, verbal interchanges include making contracts, emotional expression, delivering rewards, etc. For more on this broad conception of what is verbal, see Skinner (1957).

including behavioral interpretations of terms such as resistance, repression, and free association. Following this work, very little was published on CBA other than the insightful, but largely ignored papers by Charles Ferster (1967, 1972a, 1972b, 1972c, 1979). Neither Ferster nor Skinner intended to devise new approaches to treatment in these writings. Instead they wrote in behavioral language, demonstrating a more useful way of describing, understanding, and in Ferster's case, teaching the change process. So, CBA got its start quite early in 1953 and then all but disappeared until its reemergence in 1987 with the publication of an edited book (Jacobson, 1987). The Jacobson book contained chapters by Hayes (1987) and Kohlenberg & Tsai (1987) that described in detail their approaches to using Skinnerian principles to treat outpatient adults. We will refer to this hiatus as the quiescent period of CBA. The reasons that behavior analysts did not pursue CBA play a role in understanding the nature of its renaissance.

### CBA's Quiescent Period

Our historical account of CBA involves the interplay of three strands involving clinical applications of behaviorism. These are behavior therapy, applied behavior analysis, and the development of the Association for the Advancement of Behavior Therapy (AABT).

Behavior therapy is the application of laboratory-based principles of learning (in the early years) to human problems and a commitment to empiricism in evaluating the effects of the treatment. Applied behavior analysis is more narrowly defined as the application of operant conditioning laboratory principles to treating and solving human problems - in other words, the Skinnerian based treatment approach. Although both are applications of Skinnerian operant principles to real human problems, CBA is distinct from applied behavior analysis in that it focuses on outpatient "talk" therapy, whereas applied behavior analysts pay very little attention to such therapy. AABT is the dominant professional/scientific organization to which behavior therapists belong and, with its journal, *Behavior Therapy*, is the primary voice of behavior therapy.

In 1966, during CBA's quiescent period, AABT was established, and its journal, *Behavior Therapy*, came into existence in 1969. Although CBA was quiescent, applied behavior analysis was not. In fact, applied behavior analysis played a very significant role in the development of behavior therapy during these early years. Applied behavior analysis was considered one of the two pillars of behavior therapy, the other being desensitization and classical conditioning based treatments.

During this period, behavior analysis was in the mainstream of behavior therapy. There was a virtual explosion of research on behavior change techniques based upon operant principles (e.g., Ayllon & Azrin, 1965; O'Leary & Becker, 1967; Wolf, Risley, & Mees, 1964). In the years 1970 to 1978, a casual tabulation of the papers published in *Behavior Therapy* showed that about 40% of the empirical and treatment papers referred to the operant terms *contingency, reinforcement, extinction, or discriminative stimulus*. Many of the published graphs were cumulative records (a favorite of behavior analysts) that showed a baseline condition, a reinforcement condition, and an extinction condition. These graphs showed how the therapist's within-session actions (e.g., applications of reinforcement and punishment, shaping, exposure to feared stimuli) produced

behavior change. It is important to point out that in these papers, the behavior changes that were the goal of treatment were also observed during the session; we will elaborate on the significance of this later. There were, of course, equivalent numbers of papers on desensitization and classical conditioning applications and it was not unusual for individual papers to have references to both.

Given that applied behavior analysis was a foundation of behavior therapy and had a strong presence both in AABT and in its journal, it might well have continued playing a substantial role. However in the 1980s and 1990s, through an unexpected, curious turn of events, applied behavior analysis became a minor presence in the pages of *Behavior Therapy* (with the notable exception discussed later in this paper) though AABT, with over 4000 members, had grown and prospered. Most behavior analysts now belong to the Association for Behavior Analysis and publish their work in another journal, the *Journal of Applied Behavior Analysis*. Our explanation for this turn of events is closely related to the quiescence and eventual renaissance of CBA.

First, behavior therapists became increasingly interested in working with adults in the outpatient psychotherapy environment. Applied behavior analysts, on the other hand, mainly worked in settings that differed from the typical psychotherapy office. Further, the kinds of problems that applied behavior analysts dealt with were not typical problems of the adult outpatient such as depression, problems of the self, difficulties in intimate relationships, and existential anxiety. Instead the behavior analyst was extremely effective in treating problems such as head banging, poor math performance, hyperactive school children, tics, mutism, towel hoarding, and lack of rudimentary self care skills in hospitalized patients with schizophrenia. So, given the growing interest in adult outpatient problems and the seeming inappropriateness of applied behavior analysis, behavior therapists became less interested in applied behavior analysis. Even more telling, many applied behavior analysts left the fold and turned to cognitive therapy for guidance in doing office-

based treatment (Hawkins, Kashden, Hansen & Sadd, 1992).

The abandonment of the Skinnerian approach in mainstream behavior therapy was based on an unfortunate and misguided assumption. The process was as follows: First it was correctly concluded that applied behavior analysis was effective for a wide variety of problems ranging from self destructive behaviors in severely disturbed children to problematic learning difficulties of children in classrooms. Second it was clear that the kinds of problems that applied behavior analysis was used for all had one thing in common, namely the focus on actual within-session occurrences of the client's problematic behavior and improvement. That is, in order to use operant techniques, the therapist had to observe the problematic behavior directly, deliver the rewards and punishments, and actually see the behavior change. Third, it was assumed that most of the problems presented by outpatient adults, such as "difficulties in intimate relationships" or "depression," or "anger" occurred only outside the therapy session in their daily life and could not be observed and reinforced directly by the therapist during typical office treatment. This last assumption, we believe, was erroneous.

In addition to this erroneous conclusion, there were other barriers that deterred behavior therapists from using Skinnerian based methods for outpatient adults. Some applied behavior analysts themselves uncritically accepted the erroneous conclusion and, as discussed by Hawkins et al. (1992), became cognitive therapists. Others inappropriately used operant techniques with adults that further added to the prevailing misconception that behavior analysis had little to offer in the adult treatment arena. For example there were procedures such as contracting, e.g., "I'll fine you a nickel if you don't do this and that kind of behavior or if you weren't nice to your wife", or only paying attention to a client with depression if they were smiling, or asking a husband to earn points for taking out the garbage that could be redeemable for sex with his wife. So, during the early years of behavior therapy and continuing till just recently, applied behavior analysts were not very

effective in devising treatments that addressed the daily life problems of adult outpatients.

Thus we agree with Wulfert, in this series, that that particular form of the Skinnerian approach failed to meet an important need for expanding the scope of behavior therapy. We also agree that this failure set the stage for the so-called "cognitive revolution," because cognitive therapy easily lent itself to office treatment.

### The Renaissance

Unfortunately, the fact that problematic client behavior actually occurs in the office was overlooked by most behavior analysts. One reason for this oversight was that behavior analysts were focusing their efforts in other settings. They were doing very well with the populations in institutions such as school classrooms, mental hospitals and even whole communities. But *Behavior Therapy* stopped publishing this kind of work because the readership had more interest in office treatment, so cognitive therapy papers appeared with increasing frequency. We suspect that those applied behavior analysts who didn't embrace the cognitive perspective, left AABT and joined the Association for Behavior Analysis, published in their own journal (*The Journal of Applied Behavior Analysis*) and more or less abdicated their role in outpatient treatment to cognitive therapists.

This situation set the stage for a renaissance of clinical behavior analysis. How did we get back on track? From our perspective, the watershed event in this whole clinical behavior analysis area was Steve Hayes' work (Hayes, 1987; Zettle & Hayes, 1982;). He took Skinner's *Verbal Behavior* (1957) and applied it to outpatient individuals while investigating the underlying principles with corresponding laboratory work. For the first author, Kohlenberg, Hayes' work was an eye opener, because early in his own career during the 1970s, he was a radical behaviorist at heart, but was unable to use Skinnerian principles when doing outpatient treatment. At that time, it was difficult to conceptualize outpatient treatment in Skinnerian terms because the framework wasn't



there. Then Steve Hayes' work changed that. There were many other contributors who added to this, but it was a concentrated effort on his part that made the critical difference. This symposium that is gathered here today is a direct result of Hayes' application of radical behavioral principles to outpatient treatment. Now there is a way for behavior analysts to start talking about what goes on in a talk therapy situation.

Behavior therapists abandoned applied behavior analysis and the idea of using the Skinnerian approach, as we said, due to an erroneous conclusion that problems presented by outpatient adults do not occur in the therapy session. We do not think that the therapy environment actually differs significantly from the client's daily life environment. On the contrary, we contend that most people's daily life problems are the same kinds of problems that occur in-vivo, during office treatment. In daily life, our problems have to do with relating to other people, and therapy requires the client and therapist to relate to one another. Functionally speaking, the way you know whether a client-therapist environment is the same or different from the environment on the outside is whether or not it evokes the same kinds of problems the client reports having outside of therapy. And in fact, if therapists take a functional view of client behavior, they see that the same kinds of client problems actually do happen in the office as in daily life. That's because the therapy situation is part of daily life: it's not separate from the natural world. That being the case, the behavior analyst who's used to working with behavior as it occurs in a classroom or Skinner box, actually has the same opportunity to do something with on-going client behavior in the outpatient therapy office.

### **A Grand Theory for Behavior Therapy**

We contend that there is a need in AABT and in behavior therapy in general for what CBA has to offer. As we see it, the major problems facing the AABT membership with its current emphasis on cognitive therapy and empirically validated treatments include the lack of a coherent theoretical base that can embrace all of the techniques used by behavior therapists (Branch, 1987). During the rapid growth period

of behavior therapy there was almost no interest in theory. Now with all the behavioral procedures that were developed, a horrendous question arose, "When do you use which procedure for what kind of person?"

In fact, this is a big question for all therapies, and we behavior analysts are actually in a good position to answer it. There is nothing in a behavior analytic approach that rules out any procedure. We can do anything. We might conceptualize it in different terms than a cognitive therapist might, but basically we can embrace every treatment procedure that AABT has ever had presented at conventions or published in their journal. We can fit it into a theoretical structure and solve the problem of deciding which procedure to use. The idea that behavior analysis offers an integrative treatment approach is very compelling. There really isn't any other theory or theoretical approach that can embrace every procedure, from cognitive interventions to free association. Kohlenberg & Tsai (1994), who used CBA to embrace psychoanalytic and cognitive therapy procedures, demonstrated the integrative possibilities of this approach.

### **The Bright Future for CBA**

There has been a recent spate of reports showing that medication is better than psychosocial treatment for a variety of disorders such as depression. If medication is in fact better, then it's not a problem. But many of us think that psychosocial treatments could be better and are preferable in the long run. Now what's the solution to the problem? We must develop more powerful treatments. Again, looking at the last two 1997 issues of *Behavior Therapy*, the innovative treatments with promising futures that were mentioned the most came from the behavior analytic tradition.

We're in a position to offer something new to the field. One strategy for doing that is to build upon something that's already there. This is fairly easy to do if you look at the integrative power of behavioral analysis referred to above. As a case example, we have done a behavioral analysis of cognitive behavior therapy for depression and have come up with some very

promising improvements that should enhance efficacy. One such improvement is based on the notion that maximum change occurs when improvements in the client's behavior are reinforced as they occur, within the therapist-client relationship. For example, a client who feels isolated because s/he always needs to appear strong, competent and in control, happens to admit a fear or a weakness to the therapist. If the therapist responds honestly that s/he feels closer to the client as a result of this disclosure, that this may help the client to risk making such a disclosure with selected others outside of therapy, consequently feeling less isolated. In other words, an in-session, directly observable client behavior (admitting a weakness to the therapist) occurred and was immediately reinforced<sup>2</sup>. We have been conducting an NIMH treatment development study to find out if these in-vivo enhancements could be implemented. Although the study is not yet completed, preliminary results indicate that experienced cognitive therapists can learn how to do the enhanced treatment. As shown in Figure 1, therapists doing the enhanced treatment with clients with depression make much more use of the therapist-client relationship as an in-vivo example of their daily life problems.

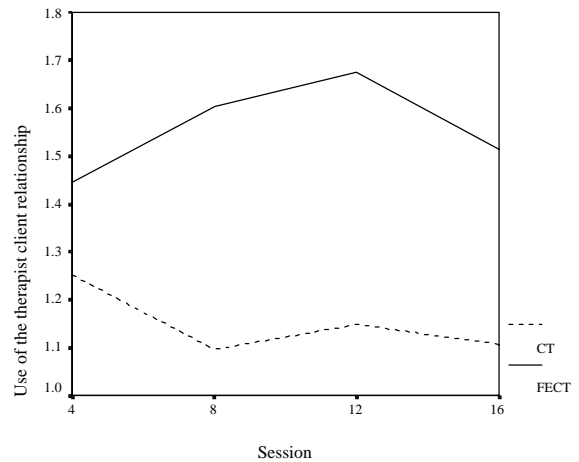


Figure 1. Use of the therapist client relationship in sessions 4, 8, 12, and 16 for clients receiving either Cognitive Therapy (CT, n=15) or the functionally-enhanced Cognitive Therapy (FECT, n=23). The use of the therapist client relationship was measured by trained raters using the Therapist In-vivo Strategies Scale (Parker, Bolling, & Kohlenberg, 1996) on videotapes of therapy sessions from the NIMH treatment development study. High scores indicate more use of the client-therapist relationship.

Behavior analysis can even help with problems that are perplexing to cognitive therapists regarding such issues as cognitive structures, cognitive products, and automatic thoughts, specifying exactly how they differ and how to change them (Kohlenberg & Tsai, 1991). We have some very good solutions to these questions that are based on distinctions between rule-governed and contingency-shaped behavior. So, not only can we improve treatment but we can also help cognitive therapy and make some friends. We agree with Wulfert that we need to reach out, offer something, and learn something, rather than just being critical of our AABT brothers and sisters. Our study on enhancing cognitive behavior therapy has shown what a good treatment cognitive therapy is and how difficult it is to do properly. We employed experienced cognitive behavior therapists and have learned to appreciate what they do. It is actually a very good treatment: it's easy to undersell cognitive therapy if you don't see it in action and appreciate how difficult it is to do properly.

Although behavior analysis was out of the picture till very recently we think the current status is good, based on certain bits of evidence. First of all, Steve Hayes is the president of AABT **and** a behavior analyst. Not only that,

<sup>2</sup> Technically, the event is not known to be reinforcing until we can observe future occurrences of the client behavior in question.

but Neil Jacobson, a recent radical behavioral convert, chided the audience in his 1991 AABT presidential address that AABT wasn't behavioral anymore and had ignored functional analysis. Secondly, if you look at AABT programs, you'll find there are more and more behavior analysts actually presenting at AABT. Third, more evidence can be found in the last two issues of the 1997 *Behavior Therapy*. These issues were devoted to an assessment of the last 30 years and the future of AABT and behavior therapy. Interestingly enough, they were edited by two behavioral analysts, Rob Hawkins and John Forsyth, good evidence that clinical behavior analysis once again has a strong presence in AABT. Some may not like the idea that we are using "influencing AABT or being in AABT" as a measure of the health of CBA. But given the size and influence of AABT, it's a meaningful measure. Fourth, if you look through those last two 1997 issues of *Behavior Therapy*, it's remarkable how much attention is being given to behavior analysis by mainstream behavior therapists. References to Kohlenberg & Tsai and Hayes are frequent. That's evidence that clinical behavior analysis is more present than it has been since the very early years. So our assessment is that the current status of CBA is good and our future is bright.

## REFERENCES

- Ayllon, T. & Azrin, N.H. (1965). The measurement and reinforcement of behavior of psychotics. *Journal of the Experimental Analysis of Behavior*, 8, 53-62.
- Branch, M.N. (1987) Behavior analysis: A conceptual and empirical base for behavior therapy. *Behavior Therapist*, 4, 79-84
- Ferster, C.B. (1967). Arbitrary and natural reinforcement. *The Psychological Record*, 22, 1-16.
- Ferster, C.B. (1972a). Clinical reinforcement. *Seminars in Psychiatry*, 4 (2), 101-111.
- Ferster, C.B. (1972b). An experimental analysis of clinical phenomena. *The Psychological Record*, 22, 1-16.
- Ferster, C.B. (1972c). Psychotherapy from the standpoint of a behaviorist. In J. D. Keehn (Ed.), *Psychopathology in animals: Research and clinical implications*. New York: Academic Press.
- Ferster, C.B. (1979). A laboratory model of psychotherapy. In P. Sjoden (Ed.), *Trends in behavior therapy*. New York: Academic Press.
- Hawkins, R.P., Kashden, J., Hansen, D.J., & Sadd, D.L. (1992). The increasing reference to "cognitive" variables in behavior therapy: A 20-year empirical analysis. *The Behavior Therapist*, 15, 115-118.
- Hayes, S.C. (1987). A contextual approach to therapeutic change. In N. S. Jacobson (Ed.), *Psychotherapists in clinical practice: Cognitive and behavioral perspectives* (pp. 327-387). New York: Guilford Press.
- Jacobson, N. S. (Ed.). (1987). *Psychotherapists in clinical practice: Cognitive and behavioral perspectives*. New York: Guilford Press.
- Kohlenberg, R.J., & Tsai, M. (1987). Functional analytic psychotherapy. In N. S. Jacobson (Ed.), *Psychotherapists in clinical practice: Cognitive and behavioral perspectives* (pp. 388-443). New York: Guilford Press.
- Kohlenberg, R.J., & Tsai, M. (1991). *Functional analytic psychotherapy: A guide for creating intense and curative therapeutic relationships*. New York: Plenum.
- Kohlenberg, R.J., & Tsai, M. (1994). Functional analytic psychotherapy: A behavioral approach to treatment and integration. *Journal of Psychotherapy Integration*, 4, 175-201.
- Kohlenberg, R.J., Tsai, M., & Dougher, M.J. (1993). The dimensions of clinical behavior analysis. *The Behavior Analyst*, 16, 271-282.
- O'Leary, K.D., & Becker, W.C. (1967). Behavior modification of an adjustment class: A token reinforcement program. *Exceptional Children*, 33, 637-642.
- Parker, C.R., Bolling, M.Y., & Kohlenberg, R.J. (1996). *Therapist in-vivo strategy scale*. Unpublished manuscript. Seattle: University of Washington.
- Skinner, B.F. (1953). *Science and human behavior*. New York: Macmillan.
- Skinner, B.F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.
- Skinner, B.F. (1974). *About behaviorism*. New York: Knopf.
- Wolf, M.M., Risley, T., & Mees, H.L. (1964). Application of operant conditioning procedures to the behavior problems of an autistic child. *Behaviour Research and Therapy*, 1, 305-312.
- Zettle, R.D., & Hayes, S.C. (1982). Rule governed behavior: A potential theoretical framework for cognitive-behavioral therapy. In P. C. Kendall (Ed.), *Advances in cognitive behavioral research and therapy* (Vol. 1). New York: Academic Press.

**Author Note:** Address editorial correspondence to: Robert J. Kohlenberg, Director of Clinical Training, Department of Psychology 351635, University of Washington, Seattle, WA. 98195-1635. Email: [fap@u.washington.edu](mailto:fap@u.washington.edu)

## CAN CONTEXTUAL THERAPIES SAVE CLINICAL BEHAVIOR ANALYSIS?

Edelgard Wulfert

University at Albany, State University of New York

Behavior therapy has become separated from its philosophical roots in radical behaviorism. In many respects, operant techniques have gradually been substituted first by cognitive-behavioral and later by cognitive interventions. The main reason for the diminishing influence of behavior analysis in the area of clinical psychology is that radical behaviorists have not paid close attention to the importance of verbal processes in psychotherapy. Through new developments in basic behavior analysis, specifically in the areas of rule-governed behavior and stimulus equivalence, contemporary clinical behavior analysts have begun to re-conceptualize behavioral interventions. They no longer assume that private events need not be taken into account as there is ample evidence that verbally constructed events introduce additional variables that interact with, and often alter, the effects of contingencies. This thinking has led to the development of new contextual psychotherapies. Research is needed to demonstrate whether these behavioral interventions present alternatives to empirically validated non-behavioral treatments. The contributions of the successful therapist to clinical outcomes also need to be clarified.

---

During the early 1980's, I was a doctoral student in clinical psychology at the University of North Carolina at Greensboro. From the beginning I was drawn to a group of graduate students and faculty with a decidedly behavioral bent, including Steve Hayes, Rick Shull, and the late Aaron Brownstein. Our group thrived on conceptual debates and we particularly wrestled with the implications of Skinner's philosophical position for our clinical work. However, despite creative case formulations and penetrating conceptual analyses, those of us who worked in the clinical trenches often felt a sense of disillusionment with behavior analysis because its methods were not as useful with adult outpatients as our behavior modification textbooks would have us believe. Operant techniques had proved most successful in relatively closed behavioral systems with institutionalized clients and children. They were far less effective with adult outpatients who possessed sophisticated verbal repertoires and came to us from complex environments to whom we had no access. Although we were well trained in a range of empirically supported behavioral techniques such as self-monitoring, skills training, systematic desensitization, and exposure-based procedures, they appeared of limited value. These techniques worked reasonably well with circumscribed complaints, especially anxiety related ones such as phobias and obsessive-compulsive behaviors. But they were much less effective for clients who presented with long-standing, diffuse maladjustments characterized by emotional

lability, low self-esteem, chronic relationship problems, and other self-destructive behaviors. For some of these clients it was difficult even to specify target behaviors; other clients were plainly uncooperative with operant procedures or they complied and did not improve; and still others improved but relapsed soon after treatment was completed. I should note that negative outcomes were not indicative of incompetent student therapists or deficient clinical training in Greensboro; rather, treatment failures in behavior therapy are a well-established fact (see Mays & Franks, 1985).

### Behavior Therapy with a Cognitive Twist

Many of us who faithfully applied operant principles in our clinical work struggled with an increasing sense of frustration and disillusionment because our field seemed to stagnate. At the same time, in behavior therapy at large a "cognitive revolution" was in full progress. A new breed of *cognitive behavior therapists* adopted a constructivist philosophy, advocating that maladaptive behavior was not caused by the external contingencies but by a dysfunctional belief system through which clients filter and interpret the events of their lives (e.g., Beck, 1976; Ellis & Dryden, 1987). Based on this premise, cognitive therapists persuaded clients that their emotional difficulties were largely self-generated by absolutistic, self-defeating beliefs. They taught clients to become aware of maladaptive self-verbalizations, identify them as members of specific categories of cognitive distortions and replace them with

more adaptive self-statements. They also trained clients to "become scientists" and empirically disconfirm their beliefs. Replacing cognitive distortions with accurate self-statements was the presumed mechanism by which clients were able to change their maladaptive behavior and lead productive lives. Concerned about the lack of innovative treatments for mainstream clinical disorders, I predicted that without a search for new horizons clinical behavior analysis would soon become obsolete. Steve Hayes, whose faith in the supremacy of behavior analysis continued unabated, predicted that over the course of the next ten years radical behaviorists would "make a contribution that will alter the course of clinical psychology." He challenged me to a bet, but as with most things in life that can be viewed from different perspectives, years later we could never agree on who won.

### Thinking and Feeling

During the 1970s and 80s, cognitive behavior therapy steadily gained ground. Despite its empirically supported clinical successes, our group of 'radical behaviorists' rejected the hybrid position. We repudiated its basic tenets, not because behavior analysts deny the existence of private experience or prohibit their scientific analysis, as opponents have often charged (e.g., Mahoney, 1989), but on pragmatic and philosophical grounds. As thoughts and feelings cannot be changed without changing their context, for practical reasons behavior analysts reserve causality for manipulable environmental events (Skinner, 1974). From a philosophical viewpoint, nothing prohibits the analysis of private experience. Although early metaphysical behaviorists (e.g., John B. Watson) in fact ruled out private events of the bounds of science, for Skinner (1945) thinking and feeling were legitimate objects of scientific study. Although behaviorists strongly value observational methods, what is observed need not be public. Given that the validity of an observation depends on the controlling variables, observing private events can be as scientifically valid as observing public behavior if the report is under tight control of the relevant stimuli.

Skinner's (1974) dictum notwithstanding, his theoretical position did not

encourage the systematic analysis of private events. Although he laid the groundwork for the analysis of verbal behavior, there was no compelling need for behavior analysts to study private events. On the one hand, applied behavior analysts successfully worked with a range of severe behavior problems including self-injurious behavior and learning disabilities without addressing private events. On the other hand, if thinking and feeling are private responses to the same contingencies that control overt behavior, it is unnecessary to study them because public behavior can be understood without them. To illustrate, if a client with agoraphobia avoids places and we assume that her private emotion 'panic' is controlled by the same contingencies, focusing on feelings does not help us to understand the avoidance. Exposing the client to feared situations and preventing escape should result both in a lessening of her avoidance behavior and her anxiety.

This example illustrates why behavior analysts traditionally have not been concerned with thinking and feeling. Skinner made no qualitative distinction between public and private responses, subsuming thinking, feeling, and overt responding all under the term 'behavior.' However, it can easily be seen that private events in this system can be relegated to the status of epiphenomena, that is, contingency-determined private correlates of public behavior that have no bearing whatsoever on its course.

### The Evolution of Contemporary Clinical Behavior Analysis

In the role of clinicians, our central goal is to help clients to change maladaptive behaviors so that they can lead personally more satisfying lives. When working with adult outpatients, this goal can be attained only indirectly through verbal interchanges in the therapist's office. What the therapist says in session becomes the main vehicle for changing a client's belief system and behavior in everyday life. This underscores how enormously important the verbal transactions between client and therapist are. At the same time it highlights how unfortunate it was that behavior analysts did not strive to better understand the mechanism by which what is said in session

affects client behavior outside of the session. Two unanticipated findings from basic operant research were instrumental in changing course and setting the occasion for the rise of a modern version of clinical behavior analysts.

First, laboratory studies on instructional control provided evidence that verbal antecedents (“rules”) tend to override the effects of nonverbal contingencies, making instructed performances, compared to shaped performances, insensitive to programmed contingencies. (For a synopsis of rule-governed behavior see Catania, 1998). This finding had practical implications. One, as telling clients what to say or do is often *not* followed by corresponding actions, shaping their verbal behavior to help them find their own solutions to their problems may lead to better results. Two, a careful analysis of clients’ rule repertoires is imperative. Understanding the rules clients learned early in life may explain why they persist in difficult situations, seemingly against reason and unable to change.

A second source of influence came from the literature on stimulus equivalence (Sidman, 1994). When people are taught specific relations between stimuli that involve verbal events, often additional relations emerge spontaneously. Once a network of relations has emerged, functions acquired by one stimulus may transfer without training to other members. Hayes and Toarmino have nicely illustrated the theoretical and practical implications of this finding. They state that clients often avoid thinking or talking about painful memories. If verbal behavior were the strictly unidirectional process implied in Skinner's view (i.e., private and public behavior being synchronized by the same contingencies), this avoidance behavior would be difficult to explain in terms of conditioning principles, as the following example illustrates. An animal can be taught to “report” that it was shocked when the experimenter reinforces one response following shock, and a different response following no shock. Shock and report are uni-directionally related. The shock is aversive and elicits pain and fear, but the ‘report’ does not because it comes *after* the shock. In contrast, when human beings think or talk about a painful experience, their verbal behavior can itself be painful

because the event and the verbal report are bi-directionally related. Reporting a private event can evoke some of the original reactions the person experienced when the event occurred, but it can also change the reactions originally produced by that event, which may be the mechanism behind insight-oriented and humanistic therapies (Hayes & Toarmino, 1999).

Stimulated by these developments, a group of clinical behavior analysts, under the leadership of Steve Hayes, began to re-conceptualize behavioral interventions. It was no longer assumed that private events need not be taken into account as there is ample evidence that verbally constructed events introduce additional variables that interact with, and often alter, the effects of contingencies. During the 1990s, several ‘contextual psychotherapies’ have been developed that are conceptually and philosophically rooted in radical behaviorism (see Hayes, Jacobson, Follette, & Dougher, 1994). The two prime examples of this modern-day clinical behavior analysis are Steve Hayes’ Acceptance and Commitment Therapy (ACT; Hayes, Strosahl, & Wilson, 1999) and Bob Kohlenberg’s Functional Analytic Psychotherapy (FAP; Kohlenberg & Tsai, 1991). Both therapies start with the premise that the verbal exchanges between client and therapist in the office are the main vehicle of change and both are suited for adult outpatients presenting with the range of disorders seen in general outpatient practice.

It would go beyond the scope of this paper to describe these contextual therapies in detail. In essence, the ACT therapist views psychological disorders largely as behaviors controlled by the effects of language. By using metaphors, paradox, and experiential exercises, the ACT therapist undermines destructive rule- and self-rule following so that clients commit themselves to behavior change in spite of what they think or feel. In FAP, verbal processes are also paramount. The FAP therapist identifies clinically relevant behaviors (thoughts and feelings the client originally experienced with significant others and now generalizes to the therapist). Maladaptive verbal repertoires are modified in session through verbal shaping techniques and the use of natural as opposed to arbitrary reinforcement (see Ferster, 1967).

Through the recent developments in basic human operant research the theoretical scope of radical behaviorism has been extended in substantive ways. With rule-governance and stimulus equivalence, clinical behavior analysts are now able to provide a theoretically coherent rationale of the purported mechanisms by which "talk therapy" affects clients' functioning in their everyday lives.

### **Can the Contextual Therapies Reclaim a Niche for Behavior Analysis in Clinical Psychology?**

If one were to conduct a careful examination of the verbal interactions between client and therapist in ACT or FAP versus psychoanalytic or existential therapy, I suspect striking similarities would be revealed. Indeed, Hayes and Toarmino (1999) as well as Kohlenberg and Tsai (1991) have freely admitted that their therapies in technical features resemble some non-behavioral approaches, although philosophically and theoretically they are grounded in behaviorism. It might seem surprising that non-behavioral techniques have found their way into clinical behavior analysis. However, we must remember that techniques, like the outcomes they generate, can be interpreted within different theoretical frameworks. This might be one of the reasons that the new contextual behavior therapies are beginning to receive attention not only from clinical behavior analysts but also from psychologists of different theoretical persuasions. To illustrate, in recent convention programs of the Association for Advancement of Behavior Therapy 'acceptance' has secured its own heading in the subject index and references to ACT and FAP are growing in the published literature. Does this mean that clinical behavior analysis is conquering mainstream psychotherapy? Will cognitive and psychodynamic psychologists, who have recently expressed dissatisfaction with the adequacy of their own theoretical models of change (Castonguay & Goldfried, 1994), embrace behavior-analytic principles? I think not.

Hayes and Toarmino (1999) as well as Kohlenberg and colleagues (this issue) assert that behavior analysis can provide the empirical-theoretical basis for traditional clinical

approaches. My colleagues' exuberance notwithstanding, I strongly doubt that non-behavioral scientists will look to behaviorists for theoretical guidance or that they can be persuaded by our data. Theories do not evolve from facts; rather, they are constructed together with the facts (Kuhn, 1996). The reason that clinical behavior analysts have begun to integrate some of the more traditional psychotherapy techniques into their interventions is that they have seen these techniques work in other contexts and they can now accommodate them conceptually in their expanded theoretical framework. Similarly, non-behavioral clinicians have used and will continue to employ certain behavioral strategies (e.g., exposure) as long as they can justify their use within their own theories. However, borrowing techniques motivates neither behaviorists nor traditional psychologists to switch paradigms. Dollard and Miller's (1950) behavioral reinterpretation of Freudian core concepts is a classic case in point: their book inspired behaviorists, but it did nothing to change the course of psychoanalysis.

I doubt very seriously that we as clinical behavior analysts will regain influence in outpatient therapy by attempting to persuade non-behaviorists of the merits of our epistemology; rather, we should reach out to colleagues of other theoretical orientations to forge an alliance based on mutual respect for our differences (Wulfert, 1997). Behavior analysts have always prided themselves in the empirical foundations of their science. In keeping with this tradition, we should work diligently to establish the efficacy and effectiveness of the new contextual therapies. This would require that these therapies be submitted to rigorously controlled randomized clinical trials and that other laboratories independently replicate these trials. Overzealous claims to the contrary notwithstanding, to date the empirical basis of both ACT and FAP continues to be weak. But I am extremely hopeful that the new behavioral interventions will indeed prove effective, especially compared to wait list or placebo controls. A more important question, however, is whether they will turn out to be *more* effective than non-behavioral treatments. Considering that comparative treatment outcome research has consistently shown that a range of empirically

validated cognitive, humanistic, and psychodynamic therapies work equally well (e.g., the NIMH Depression Study or Project MATCH) I would not be surprised if the new behavior-analytic ‘poster-children’ would fare no better in rigorous comparisons against established non-behavioral treatments. What would this mean?

One reasonable implication of these findings is that psychotherapy clients improve not because of our “theory-based interventions” but because of the effects of individual psychotherapists. Recent research seems to suggest this quite strongly (e.g., Lambert & Okiishi, 1997; Luborsky et al., 1986; Luborsky, McLellan, Diguier, Woody, & Seligman, 1997). One of our own, Charles Ferster (1972), over a quarter of a century ago chided behavior analysts to pay more attention to what effective therapists do rather than quibbling about the merits of their theoretical explanations. If it turned out that most of the variance in psychotherapy outcome is indeed attributable to the therapist, perhaps we should have listened?

## REFERENCES

- Beck, A.T. (1976). *Cognitive therapy and the emotional disorders*. New York: New American Library.
- Catania, C. A. (1998). *Learning* (4<sup>th</sup> ed.), pp. 264-270. Englewood Cliffs, N.J.: Prentice Hall.
- Castonguay, L. G., & Goldfried, M. R. (1994). Psychotherapy integration: An idea whose time has come. *Applied and Preventive Psychology*, 3, 159-172.
- Dollard, J., & Miller, N.E. (1950). *Personality and psychotherapy*. New York: McGraw-Hill Book Company.
- Ellis, A., & Dryden, W. (1987). *The practice of rational-emotive therapy (RET)*. New York: Springer.
- Ferster, C. B. (1967). Arbitrary and natural reinforcement. *The Psychological Record*, 22, 1-16.
- Ferster, C. B. (1972). An experimental analysis of clinical phenomena. *The Psychological Record*, 22, 1-16.
- Hayes, S. C., Jacobson, V. M., Follette, V. M., & Dougher, M. J. (Eds.) (1994). *Acceptance and change: Content and context in psychotherapy* (pp. 203-217). Reno, NV: Context Press
- Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (1999). *Acceptance and commitment therapy*. New York: Guilford Press.
- Hayes, S. C., & Toarmino, D. (1999). The rise of clinical behaviour analysis. *The Psychologist*, 12, 505-509.
- Kohlenberg, R. J., & Tsai, M. T. (1991). *Functional analytic psychotherapy: Creating intense and curative therapeutic relationships*. New York: Plenum Press.
- Kuhn, T. S. (1996). *The structure of scientific revolutions* (3<sup>rd</sup> ed.). Chicago: University of Chicago Press.
- Lambert, M. J., & Okiishi, J.C. (1997). The effects of the individual psychotherapist and implications for future research. *Clinical Psychology: Science and Practice*, 4, 66-75.
- Luborsky, L., Crits-Christoph, P., McLellan, A.T., Woody, G., Piper, W., Imber, S., & Liberman, B. (1986). Do therapists vary much in their success? Findings from four outcome studies. *American Journal of Orthopsychiatry*, 56, 501-512.
- Luborsky, L., McLellan, A. T., Diguier, L., Woody, G., & Seligman, D. A. (1997). The psychotherapist matters: Comparison of outcomes across twenty-two therapists and seven patient samples. *Clinical Psychology: Science & Practice*, 4 (1), 53-65.
- Mahoney, M. J. (1989). Scientific psychology and radical behaviorism: Important distinctions based in scientism and objectivism. *American Psychologist*, 44, 1372-1377.
- Mays, D. T., & Franks, C. M. (Eds.) (1985). *Negative outcome in psychotherapy and what to do about it*. New York: Springer.
- Sidman, M. (1994). *Equivalence relations and behavior: A research story*. Boston: Authors' Cooperative.
- Skinner, B. F. (1945). The operational analysis of psychological terms. *Psychological Record*, 52, 270-276.
- Skinner, B. F. (1974). *About behaviorism*. New York: Knopf.
- Wulfert, E. (1997). The exodus of behavior analysis: Is “splendid isolation” the way to go? *Behavior and Social Issues*, 7 (2), 107-112.

**Author Note** Address editorial correspondence to: Edelgard Wulfert, Ph.D., University at Albany, State University of New York, Department of Psychology, 1400 Washington Ave., Albany, NY 12222. Email: [ew302@cnsibm.albany.edu](mailto:ew302@cnsibm.albany.edu)



## SOCRATES AND THE DODO BIRD: CLINICAL BEHAVIOR ANALYSIS AND PSYCHOTHERAPY RESEARCH

Elizabeth V. Gifford  
University of Nevada, Reno

Clinical behavior analysis (CBA) is characterized by inductive observation and a pragmatic focus on outcomes. This paper argues that CBA may be useful in addressing certain issues confronted by the field of psychotherapy research. These issues include the Dodo bird dilemma, where many therapies with different purported mechanisms appear to have the same results; and Gordon Paul's famous maxim, which requires sorting out the process by which different therapies affect different individuals. A behavioral analysis of experiential avoidance is presented and discussed as one example of a behavior analytic approach to therapy process.

---

One of the strengths of Clinical Behavior Analysis (CBA) is its humble posture toward its subject matter. As Skinner wrote, "the subjects [are] always right. They always behave as they ought" (Skinner, 1972, p. 123). The truism that in order to hear one must first be willing to listen applies to this position. In order to listen carefully to what our subjects have to tell us, to be as receptive as possible to the events of interest, clinical behavior analysts consciously attempt a posture of ignorance. The inductive tradition underlying CBA requires that we stop talking/theorizing, and force ourselves to be quiet and pay attention. This Socratic respect for ignorance is an important part of what we do and what we have to offer to the field as a whole: careful, honest observation.

A second and related strength of CBA is its commitment to a pragmatic standard. As behaviorists, our interest is the prediction and influence of behavior. This means that we are not only careful observers as we enter a situation, we are also careful observers of the effects of our participation<sup>1</sup>. The methods of CBA hold us accountable in that our interventions must be linked to meaningful changes in our client's actions. And in the broader field of clinical psychology, a field historically complicated by opinion, myth and other cultural heuristics, a commitment to accountability is a significant strength.

So what might these principles look like, applied to the field of psychotherapy research? Certainly there are some interesting questions

facing this field at the moment. For example, large multi-site clinical trials comparing different psychotherapies, such as the NIMH Treatment of Depression Collaborative Study (Elkin, Parloff, Hadley, & Autry, 1985) and Project MATCH (Project MATCH Research Group, 1997), appear to result in moderate effect sizes of roughly the same magnitude. So in some areas we are confronted by the "Dodo bird verdict," where in a Lewis Carroll world everyone urges on their favorite therapy while the Dodo bird happily declares that "everyone has won and all must have prizes!" (Luborsky, Singer, & Luborsky, 1975, p. 995). Thus, current challenges in psychotherapy research include the following: (a) the effects of psychotherapies are often equivalent, even psychotherapies with ostensibly very different ingredients, and (b) in these cases the critical processes involved in clinical change remain unclear.

Similar concerns prompted Gordon Paul's famous question: "*what* treatment, by *whom*, is most effective for *this* individual with *that* specific problem, and under *which* set of circumstances?" (Paul, 1967, p. 111). There have been a number of solutions offered to this dilemma. One solution has been to gather variables within categories and create a matrix of treatment X therapist X client X problem X setting factors. A problem with this strategy is that it quickly leads to an inordinate number of cells. For example, evaluating 10 therapists X 10 clients X 10 treatments X 10 problems X 10 settings would require a matrix of 10,000 cells (Stiles, Shapiro, & Elliot, 1986). Such solutions could keep the Dodo bird happy for the indefinite future.

<sup>1</sup> William James states: "An attitude of orientation, is what the pragmatic method means. The attitude of looking away from first things, principles, "categories," supposed necessities; and of looking towards last things, fruits, consequences, facts" (James, 1995, p. 57).

In contrast, CBA encourages us to go back and ask the data where and how to begin. What are the active processes that occur in sessions? Which processes are critical in successful/unsuccessful treatments? Are there important processes that occur across therapies? Indeed, Gordon Paul later expanded his question to include “*by what process?*” and emphasized that this was the most critical component of his formulation (Paul, 1999). Stated differently, in order to understand our successes and our failures we need to identify the mediating variables influenced by our interventions (Follette, 1995). Mediating variables explain how or by what means a predictor variable effects an outcome variable. For example, do exposure and response prevention treatments lose effectiveness when clients successfully distract themselves during treatment (Grayson, Foa & Steketee, 1982)? The critical factor is not whether a therapeutic technique is applied; it the occurrence of the variables/processes that are related to clinical change.

There are any number of phenomena we might observe if we are willing to let the raw data of behavior guide us. How are we to organize and make sense of the extraordinarily complex worlds of psychopathology and psychotherapy? CBA offers basic behavior analytic principles, which are patterns of relationships (e.g., operant and respondent relations) that have been abstracted inductively from numerous observations and sets of observations. These basic principles can, in turn, aid in identifying meaningful patterns in complex phenomena.

As an example I would briefly like to discuss experiential avoidance, defined as maladaptive efforts to alter the form or frequency of certain thoughts, feelings, and/or bodily sensations (Hayes, Wilson, Gifford, Follette & Strosahl, 1996). Remarkably enough, most systems of therapy recognize experiential avoidance as a pathogenic process, across a wide variety of disorders (see Hayes et al., for an extended discussion). Additionally, research in the experimental laboratory supports the hypothesis that efforts to avoid or control thoughts and feelings may be problematic (see Wegner, 1994). For example, analogue research examining the suppression of thoughts with

traumatic content has found that efforts to suppress such thoughts increase traumatic intrusions. Such avoidance or suppression processes are hypothesized to contribute to the perpetuation of post-traumatic stress symptoms experienced by certain individuals after exposure to a traumatic event (Davies & Clark, 1998; Harvey & Bryant, 1998). This convergence of observations on experiential avoidance - across both theoretical schools and basic and applied settings - supports further investigation.

Applying behavioral principles to this phenomenon, one hypothesis is that experiential avoidance functions as a negatively reinforced behavior. For the person with alcoholism who drinks to avoid feeling negative feelings - and up to 80% of alcoholics' relapse episodes may be aimed at this purpose - the immediate benefits of drinking are obvious (Wulfert, 1994). Once behavior is under the control of these immediate reinforcers, suppression and other processes may increase the frequency of the aversive experience, contributing to self-perpetuating avoidance cycles that can prove difficult to alter (Hayes & Gifford, 1997).

As I have discussed here previously (Gifford, 1997), this view of experiential avoidance has certain implications. One implication is that acceptance interventions should begin by acknowledging the impossibility of completely or permanently preventing the occurrence of certain thoughts and feelings. Who could imagine a world where clients will not in the future be challenged by difficult emotions or cognitions? Focused training in acceptance should then be directed at altering the client's responses in the presence of the previously avoided experience(s). Exposure to the dysphoric mood or troublesome cognitions while emitting an alternative response permits contact with alternative reinforcers (such as meaningful interpersonal interactions, meeting personal goals, etc.). As these alternative responses are reinforced in the presence of the previously avoided experience, the previously avoided experiences acquire approach or acceptance functions. Such functions do not change the stimuli themselves, including their negatively valenced qualities; they do, however, add to response options

available in their presence. Experiential stimuli previously associated with maladaptive behavior (e.g., smoking, drinking, etc.) develop relations to other repertoires. A final implication is that these changes should be practiced across multiple exemplars, in order to enhance generalization (Goldstein, 1995).

Based on the above analysis, acceptance-oriented therapies may want to include the following:

1. A therapeutic context that establishes motivational social reinforcement, interpersonal modeling of experiential acceptance, and direct shaping of acceptance behavior.
2. Cognitive, affective, and experiential self-discrimination skills.
3. Guided exposure to previously avoided thoughts and feelings, with response prevention.
4. Constructive behavioral activation in both the presence and the absence of previously avoided thoughts and feelings, across multiple exemplars.

This brief analysis is just one example of applying CBA to a complex clinical phenomenon. How useful it proves remains to be seen. However, beyond any particular analysis, it is the principles by which we work that will shape our future. The clinical professional landscape is in a state of evolution, with many changing opportunities and adaptational demands. CBA's commitment to pragmatic, inductive principles may help in the quest to identify critical processes in therapeutic treatments. The goal is to give the future what it requires of us: not only those questions we can answer, but the questions we have yet to ask.

## REFERENCES

- Davies, M.L. & Clark, D.M. (1998). Thought suppression produces a rebound effect with analogue post-traumatic intrusions. *Behavior Research and Therapy*, *36*, 571-582.
- Elkin, I.E., Parloff, M.B., Hadley, S.W., & Autry, J.H., (1985). NIMH treatment of depression collaborative research program: Background and research plan. *Archives of General Psychiatry*, *42*, 305-316.
- Follette, W.C. (1995). Correcting methodological weaknesses in the knowledge base used to derive practice standards. In S.C. Hayes, V.M. Follette, R.M. Dawes, & K.E. Grady (Eds.), *Scientific standards of psychological practice: Issues and recommendations* (pp. 229-247). Reno, NV: Context Press.
- Gifford, E. V. (May, 1997). Discrimination training and the function of acceptance. Paper presented at the meeting of the Association for Behavior Analysis, Chicago, IL.
- Goldstein, A.P. (1995). Coordinated multitargeted skills training: The promotion of generalization enhancement. In W. O'Donohue & L. Krasner (Eds.), *Handbook of psychological skills training: Clinical techniques and applications*. (pp. 383-399).
- Grayson, J.B., Foa, E.B., & Steketee, G. (1982). Habituation during exposure treatment: Distraction vs. attention-focusing. *Behaviour Research and Therapy*, *20*, 323-328.
- Harvey, A.G. & Bryant, R.A. (1998). The effect of attempted thought suppression in acute stress disorder. *Behaviour Research and Therapy*, *36*(6), 583-590.
- Hayes, S.C., & Gifford, E. V. (1997). The trouble with language: Experiential avoidance, rules, and the nature of verbal events. *Psychological Science*, *8*, 170-173.
- Hayes, S.C., Wilson, K.G., Gifford, E.V., Follette, V.M., & Strosahl, K. (1996). Experiential avoidance and behavioral disorders: A functional dimensional approach to diagnosis and treatment. *Journal of Consulting and Clinical Psychology*, *64*, 1152-1168.
- Luborsky, L., Singer, B., & Luborsky, L. (1975). Comparative studies of psychotherapies: Is it true that "Everyone has won and all must have prizes"? *Archives of General Psychiatry*, *32*, 995-1008.
- Paul, G.L. (1967). Strategy of outcome research in psychotherapy. *Journal of Consulting Psychology*, *31*, 109-118.
- Paul, G.L. (1999). *The active unconscious, symptom substitution, and other things that went bump in the night*. Paper presented at the Conference on the History of Behavior Therapies, Reno, NV.
- Project Match Research Group (1997). Matching alcoholism treatments to client heterogeneity: Project MATCH posttreatment drinking outcomes. *Journal of Studies on Alcohol*, *7*-29.
- Skinner, B.F. (1972). A case history in scientific method. In B.F. Skinner (Ed.), *Cumulative record: A selection of papers* (3<sup>rd</sup> Ed.), (pp. 101-124). New York: Appleton-Century-Crofts. (Original work published 1956).
- Stiles, W.B., Shapiro, D.A., & Elliot, R. (1986). "Are all psychotherapies equivalent?" *American Psychologist*, *41*, 165-180.
- Wegner, D.M. (1994). Ironic processes of mental control. *Psychological Review*, *101*, 34-52.
- Wulfert, E. (1994). Acceptance in the treatment of alcoholism: a comparison of Alcoholics Anonymous and social learning theory. In S.C. Hayes, N.S. Jacobson, V.M. Follette, & M.J. Dougher (Eds.), *Acceptance and change: Content and context in psychotherapy* (pp. 203-217). Reno, NV: Context Press.

**Author Note:** Address editorial correspondence to: Elizabeth V. Gifford, email address: [elizgifford@earthlink.net](mailto:elizgifford@earthlink.net)

## CONSTRUCTING A BEHAVIOR ANALYTICAL HELPING PROCESS

John L. Williams  
University of Waterloo

The language of behavior analysis and radical behaviorism can provide a coherent framework for the helping processes we call counseling or psychotherapy. This can best be done when we use concepts consistent with this framework and avoid the use of mentalistic terminology that is part of psychiatric diagnostic systems and various "cognitive" schools of counseling/psychotherapy. Goldiamond's constructional model is an example of a thoroughly behavior-analytic approach that can readily be applied in the counseling setting. This model is also very compatible with and was likely an influence on current schools operating within a solution-focused orientation.

---

This piece is a digest of some of the major themes included in presentations and workshops that I have given for a number of years at the conventions of the Association for Behavior Analysis. I feel that scientists and practitioners who work within the perspective of behavior analysis and its philosophy of science, radical behaviorism, should maintain a vocabulary and practice that are consistent with this framework. I believe the following items are essential to operating within this perspective.

First, the language and semantics of the field are of critical importance. I am not sure whether the field of "clinical" behavior analysis *should* have a future. In saying this, I do not mean to assert that I think this work should stop, but that I have a problem with the term, "clinical behavior analysis". I realize, however, that it will be difficult to supplant it because of the strong tradition in psychology that is built around clinical psychology. I was trained as a clinical psychologist but I work in that somewhat distinctive environment called a counseling center and I call myself a counselor. I also avoid the use of any labels related to the so-called medical model – a model that I see as mentalistic when applied in the helping arena. I am somewhat of a purist and I try to avoid using any mentalistic terms such as "mental health" or "mental illness," and I think other behavior analysts should do likewise. While I realize that many behavior analysts work within environments where they have to deal with these constructs, I can talk without difficulty to, for example, the people in our Health Services without using terms like "diagnosis" or "treatment." I do not categorize my clientele in terms of the DSM-IV. A behavioral and developmental vocabulary provides for a comprehensive involvement and practice in the

helping world that is also more optimistic than the DSM-IV.

Further, in relation to a nonmentalistic view, I do not use the term, "cognitive," and it makes me very uncomfortable to hear this word. Within the language of behavior analysis we can, as needed, adapt and indeed subsume the counseling techniques developed under the cognitive rubric. Skinner's analysis of verbal behavior (Skinner, 1957) and his concept of rule-governed behavior, which he was using by the early 1960's (Skinner, 1964), gives us this ability. I find that when I talk in terms of rules with my clients, they typically understand this more quickly and easily than they do the word "cognition". Even when they have taken psychology courses, they do not find the latter word as useful.

Cognition is a mentalistic term that is used because of its compatibility with everyday language but also in part because it functions to enable humans to maintain, through conceptual means, their position at the top of the animal hierarchy. Further, the use of a cognitive vocabulary seems related to the culture's discomfort with the view that humans may operate pretty well as other organisms do.

Moving to a second important consideration with respect to the future of behavior analysts as helping professionals, I think that Goldiamond's (1974) article, "Toward a Constructional Approach to Social Problems," (expanded upon in a book by Swartz and Goldiamond, 1975) is a landmark in relation to behavior analytic counseling. Both this article and book have largely lain fallow for many years. The more recent advent of "solution-focused" approaches (de Shazer, 1988),

however, has opened the door for another look at Goldiamond's approach. After receiving training and reading some of the literature in the solution-focused area, I reread Goldiamond's article and saw that he anticipated all the main features of the solution-focused method. The solution-focused community has, however, developed some techniques that facilitate use of the constructional approach. In both viewpoints the focus is positive and forward moving. The work is focused on the present and toward the future. We are here to help facilitate *competence*, not mental health. We are building effective actions, not eliminating symptoms. We are also going to do this most effectively, as Goldiamond suggested, if we use people's available repertoires. Similar to programmed instruction, we best enable people to deal with the difficult situations in their lives if we use and adapt the capabilities they have already demonstrated.

A second aspect of Goldiamond's model is to place the focus of change on the everyday environment. This is consistent with the history of behavioral counseling approaches. If a client is seen for a one-hundred-and-sixty-eighth of a week, which is one hour out of every week, the counselor had best work within a context of discussion related to the client's everyday environment. This is quite different from the founding credo fostered by the psychodynamic and client-centered orientations wherein client change is seen as mainly dependent on an examination of the interchange between analyst/counselor and client.

The question of "resistance" is illustrative of the differences between the traditional and constructional/solution-focused perspectives. In the view of the latter model, resistance is not considered to be a useful concept. If a client is not collaborating, the counselor needs to discern the function(s) the client's behavior is serving. When the client's goals are ascertained, the counselor can prompt more appropriate and effective alternative behavior based on exploration of the client's fluent or available repertoires.

Given the tone of this article, the use of a medical model metaphor at this point may seem surprising, but it is a fitting one:

Goldiamond called the traditional view of the helping process the *appendix* model and contrasted it with the *pneumonia* model. Psychodynamicists and some humanists handed us the task of seeing persons over an extended period of time and producing such profound effects that these persons can then handle anything in their future lives. Once our appendix is removed, it will never again cause us problems. On the other hand, we may get pneumonia repeatedly. We have to go back to our physician occasionally to get help with pneumonia or other kinds of medical problems. Personal situations are like this; they recur and change over time. Similarly, if we see a lawyer to make a will, this does not take care of arranging a house mortgage or suing our MD if we feel our MD has mistreated us. We will go back to the same or another lawyer. Counseling best serves clients, not by trying to remake them, but by working with them on the particular and concrete goals that will help them deal with their present and prospective environments. This naturally leads to solution-focused interventions which will set clients on a more positive track that may see the solutions generalize to other circumstances. Clients may, however, come back to you after a period of time and there is no problem with that.

With the culture's steadily increasing awareness, acceptance, and use of counseling/psychotherapy, cost becomes a consideration. Only the wealthy can afford the long-term relationships that have been a part of traditional approaches.

If we look at the helping process in the constructional/solution-focused manner, this affects the kinds of data we need to collect. The clearly defined goals that are part of constructional work facilitate the assessment of effectiveness in terms of those goals rather than whether a "cure" has been obtained. One approach to getting this type of data lies in the adaptation of Goal Attainment Scaling (Kiresuk & Sherman, 1968) or similar procedures that can help articulate and individualize outcome measurement. These measures are not as exact as conventional behavior analytic methods but they do allow us to appraise our effectiveness in relation to the goals of the individual client.

In summary, several features have been delineated essential to a counseling/psychotherapy model that are consistent with a behavior analytic and radical behavioral orientation. These features include (a) a consistent and nonmentalistic terminology, (b) the constructional view as the fundamental operating framework, and (c) a focus on explicit and implicit goals related to the client's everyday environment.

## REFERENCES

de Shazer, S. (1988). *Clues: Investigating solutions in brief therapy*. New York: Norton.

Goldiamond, I. (1974). Toward a constructional approach to social problems: Ethical and constitutional issues raised by applied behavioral analysis. *Behaviorism*, 2, 1-84.

Kiresuk, T.J., & Sherman, R.E. (1968). Goal attainment scaling: A general method for evaluating comprehensive community mental health systems. *Community Mental Health Journal*, 4, 443-453.

Schwartz, A., & Goldiamond, I. (1975). *Social casework: A behavioral approach*. New York: Columbia University Press.

Skinner, B.F. (1964). Behaviorism at fifty. In T. Wann (Ed.), *Behaviorism and phenomenology* (pp. 79-108). Chicago: Chicago University Press.

Skinner, B.F. (1957). *Verbal behavior*. New York: Appleton-Century Crofts.

**Author Note:** Address editorial correspondence to: John L. Williams Ph.D., email: [jwilliam@watserv1.uwaterloo.ca](mailto:jwilliam@watserv1.uwaterloo.ca)

## AN OVERVIEW OF SOME CURRENT CHALLENGES WITHIN THE FIELD OF CLINICAL BEHAVIOR ANALYSIS

Erik M. Augustson  
National Cancer Institute

Clinical Behavior Analysis (CBA) is facing a number of difficult challenges and in many ways is at an important crossroads. This paper outlines several key issues and discusses possible directions CBA might pursue within the context of the current marketplace. Among the challenges faced today are a lack of common ground between academic psychologists and community-based psychologists and limited dialogue between CBA and mainstream clinical psychology. These are further complicated by economic strains that have led to increasing pressure to change the nature of clinical practice including the use of practice guidelines and empirical based treatment algorithms. Although promising, the basic assumptions of these types of treatment models may be premature. This paper explores some of the possible associated limitations and questions the use of empirical approaches to the exclusion of other treatment modalities at this time. Despite these challenges, CBA may be in a unique position to advance clinical psychology and is seen as having much to offer. The potential strengths of CBA are discussed and recent examples of the use of CBA within traditional clinical settings are highlighted.

---

One has only to pick up a recent copy of the American Psychological Association's *Monitor* or a similar publication to realize that clinical psychology is facing a number of changes and challenges that will greatly impact the field as we now know it. Clinical Behavioral Analysis (CBA) may have much to offer the field at this time, but there are obstacles, some from within our own history, which may prevent us from effectively participating in the upcoming evolution of clinical psychology. My various roles as a medical school faculty member are such that I have a foot firmly planted in both the fee-for-service clinical camp as well as the ivory towers of academia. Based on my experiences in trying to balance these two different sets of contingencies, I would like to identify some of the key issues impacting CBA and clinical psychology in general.

One observation is that we as academic clinicians need to start spending more time talking to community-based clinicians. There has been prominent coverage in trade publications such as the *Monitor* and discussions in recent editions in *The Behavior Therapist* (e.g. Raw, 1999a; 1999b; 1999c). However, many academic clinicians seem to continue to be unaware of or unconcerned about the growing pressures that community-based clinicians are under and the level this pressure is reaching. Part of this lack of awareness appears to be due to a widening schism between what academic clinicians are doing and what community-based clinicians are doing. Despite the ongoing

research efforts of academic psychologists, community-based clinical psychologists are paying less and less attention to that body of work. There is already a general attitude expressed by many of the community-based clinical psychologists I encounter that much of what we do as researchers is not relevant to what they do. The separation between research and clinic practice is going to get wider if we do not address this issue.

In particular, there is a need for clinical behavior analysts, who to date have largely worked in academic settings, to communicate more with other kinds of clinical psychologists. Traditionally CBA researchers have had even less dialogue with community-based clinicians than other academic psychologists. This has continued in large part due to the marked discrepancies in philosophical orientation between CBA and the cognitive-behavioral orientation of most community-based practitioners. Behavior analysis has a long history of not attending to, if not openly vilifying, the research and clinical activities of other psychological orientations, especially cognitive psychology. There are important historical reasons for this conflict which are beyond the scope of this paper. However, if CBA clinicians continue to ignore mainstream clinical psychology, faced with the changes that are coming in our field, we are going find ourselves cut out and left behind. This seems particularly ironic, given that the pragmatic stance of behavior analysis should allow for the

integration of other approaches based on their efficacy.

Perhaps as a first step toward closing the gap between CBA and community-based psychologists we should briefly consider the current context in which many, if not most, clinical psychologists are working. As most of us are aware, psychologists who make their livings primarily by providing clinical services on a fee-for-service basis have been facing increasing pressures. These have included reimbursement rates and number of allowed visits. The recently initiated test case lawsuits against Health Maintenance Organizations (HMO) by psychologists and the American Psychological Association are clear indications of the level these pressures have reached (See Raw, 1999b). In addition, community-based clinical psychologists are also starting to see similar financially motivated actions from sources that have a long history of reimbursing well for psychological services. For example, a large traditional fee-for-service third party provider in Alabama recently reduced its usual and customary rates by 20% for psychological services. As one can imagine, this represents a significant salary reduction for many clinicians.

Ironically, at the same time these economic down turns have been occurring in psychology, there has been an increase in the perceived importance of mental health and demands by consumer groups for better mental health coverage (e.g., National Institutes of Health, 1995). One of the outcomes of this movement has been the emergence of proposed parity laws within a number of states which are seeking to legislate that insurance companies provide the same levels of coverage for treatment of psychological conditions as they do for medical conditions.

One response to pressures from HMOs has been a call for more clearly defined practice guidelines (See discussion by Hayes, 1998a; 1998b) and the use of treatment algorithms and empirically based treatments (For example see Kendall & Chambless, 1998), a parallel move to that seen increasingly in many medical settings. There are a variety of possible positive benefits from this movement, but there are a number of potential problems as well. Obviously, there can

be great value in using empirical criteria to assess the efficacy of treatments and, given the pragmatic roots and sound methodological foundation of behavior analysis, CBA may have much to offer this endeavor. How to go about doing this is less clear to me and I find I have many questions but few answers.

Some of my concern is driven by the observation that many of the professionals who are the strongest advocates for defining empirically based treatments are doing so from the perspective of a classic cognitive-behavioral approach. Obviously as a behavior analyst, I am not convinced that the application of solely cognitive-behavioral interventions is necessarily the most appropriate form of treatment. In addition, as the push for empirically based treatments is at least in part economically driven, one goal of this movement would seem to implicitly be to carve out sections of the market place and remove practitioners who are using alternative approaches. Given the long-standing tension and conflict between cognitive psychology and behavior analysis, I suspect there are a number of practitioners who would welcome the demise of CBA.

Although I have many concerns, my strongest regarding the development of empirically based treatment algorithms, especially if the goal is to exclude other approaches, centers around the question of whether we actually have the appropriate technology to accurately identify the most appropriate treatment for a given problem. Put another way, I am not sure we know what a good treatment outcome looks like or specifically how we achieved it. Yes, we can measure changes in certain behaviors, but if our clients receive effective treatment they experience a phenomenon far richer than the changes we are measuring. Put simply, our current outcome measures may not adequately capture the complexity of the variables involved.

The meta-analytic studies of therapy outcome, which might be simplistically summarized as indicating that many treatment approaches can be equally effective, serve as an example of the potential limitations of our ability to discriminate between effective treatments (Smith, Glass, & Miller, 1980;



Svartberg & Stiles, 1991). Although a variety of conclusions can be formed based on these results, one interpretation is as an indication that psychology is not yet prepared to delineate the single most effective means to address a given clinical problem. This may suggest that we lack truly sufficient outcome measures or that there are aspects of effective treatment that are not captured by our current data analysis techniques. Whereas psychotherapy can impact patients' lives in many ways which can be assessed meaningfully by current outcome measures, it is also apparent that there may be fundamental processes of change which are not adequately assessed. As such, there may be gaps in our knowledge that must be filled before psychology is able to identify what kinds of treatments are likely to work and in what settings.

Given the above discussion, decisions need to be made by behavior analysts regarding where and how CBA will fit into the larger marketplace. The answers may lie in considering what CBA has to offer clinical psychology. On a fundamental level, we have the strength of a well-grounded theory and a lengthy history of effective research. We also have our historical roots in pragmatism that should allow for the adaptation and integration of techniques and approaches from other branches of clinical psychology. One example of how using a behavioral analytic approach can substantially add to our conceptualization of issues relevant to clinical psychology is the recent paper by Hayes and his colleagues (Hayes, Wilson, Gifford, Follette, & Strosahl, 1996). In this paper, Hayes et al. outline a formulation of the etiology and maintenance of many forms of psychopathology based on experiential and emotional avoidance. The authors of this paper, drawing on a wide range of basic and applied literature, build a convincing case for the ubiquitous nature of experiential avoidance and then apply their formulation to a variety of clinical phenomena.

The types of skills used by Hayes and his colleagues to perform that analysis are fundamental to behavior analysis and should place CBA in a prime position to contribute to the field of clinical psychology as a whole. However, largely, we have not made a substantial contribution. One reason for this

may be CBA's reluctance to move into settings in which our control of the contingencies has been limited. To truly have an impact on the field and the marketplace, CBA will need to move beyond traditional operant treatment settings such as those involving persons with developmental disabilities and closed communities/hospital settings. CBA must accept the challenges, and frustrations, of working in settings where the clinician has limited or no control over the contingencies outside of the therapy session. Despite Skinner's (1957) discussion and analysis, clinical behavior analysts have been slow to confront the impact of verbal behavior within the therapeutic context where such behavior is crucial to performing therapy with most individuals seeking treatment.

A number of clinical behavior analysts have recently been attempting to identify the means to begin to move CBA into more traditional clinical settings, that is those in which relatively high functioning, verbal adults serve as the clients. Attempts have been made to discuss issues relevant to these types of clinical settings (Dougher, 1993; 1994; Follette, Naugle & Callaghan, 1996; Hayes, Jacobson, Follette & Dougher, 1994) and developed interventions from a perspective that is also consistent with the underpinnings of behavior analysis (Hayes, 1995; Kohlenberg & Tsai, 1991). These efforts represent a significant move forward to involve CBA in clinical contexts where the therapist has minimal control and where attending to verbal behavior is key to effective clinical intervention. In particular, the growing literature on the concept of acceptance (Hayes et al., 1994) offers an advancement in the clinical formulation of CBA and a serious attempt for CBA to address highly relevant issues commonly found in many clinical settings with verbal adult clients.

Despite these advances, a great deal of work is still needed to refine these conceptualizations and treatments. For example, more attention might be focused on developing CBA therapies that can be effectively performed within a brief time frame. In addition, it will be important to come to a better understanding of the concept of acceptance which has been defined in different ways, and further elucidation of what role, if any, it plays in behavior change. This is an exciting area in which I hope to see

more work being done, and to that end we have recently initiated a research project at the University of Alabama at Birmingham to begin looking at the role of acceptance in chronic pain settings.

Despite my strong interest in acceptance, the development of therapies based on behavioral analytic principles, and how they fit into the larger clinical picture, it would be premature to abandon established behavioral and cognitive-behavioral techniques. Indeed, replacing such techniques may be an ineffective goal to even consider. Rather, it would seem more beneficial to integrate such approaches into our attempts to identify effective treatment and find ways to augment existing approaches. Although philosophical differences can present significant obstacles, integration between mainstream and CBA approaches may prove fruitful. As an example, in a recent case study, Paul, Marx & Orsillo (1999) coupled cognitive-behavioral and CBA techniques to create an effective treatment for a court-referred exhibitionist.

Another area to which CBA should attend regards the impact these changes in clinical practice will have on our students. Consideration should be given to the question of whether we are training students in the necessary skills to succeed in the current marketplace. Given the changes clinical psychology is undergoing, a re-evaluation of some of the elements of our training programs may be appropriate. Let me emphasize that I am not advocating a discontinuation of the academic core found in most clinical training programs, rather suggesting it would be useful to broaden the skills in which we train our students. For example, increasing student awareness of current professional issues/changes and keeping them informed about the ongoing national debates and the resolutions of those debates would seem key to producing well prepared professionals.

There are several behavioral repertoires that our students might benefit from acquiring, but are slightly outside the training offered by most programs. These include training in such areas as program evaluation, which has traditionally been more associated with Public

Health models, and actively exposing students to the fundamentals of how to provide clinical supervision. Increasingly, under managed care models of clinical practice, psychologists are placed in the position of overseeing the clinical work of Masters degree level clinicians. That can present a number of problems for CBA. The most striking of which is that Masters level clinician rarely, if ever, receive training in behavioral analytic theory or therapy. In that context, we need to find ways to effectively train and supervise a group of clinicians whose backgrounds are vastly different from ours. To this end, a recent study by Strosahl, Hayes, Bergan, and Romano (1998) highlights how this can be accomplished. They used field effectiveness techniques to assess whether or not Acceptance and Commitment Therapy (ACT) could be effectively performed by Masters level clinicians in a managed care setting. Their results demonstrate that Masters level therapists from diverse backgrounds can be trained to effectively apply CBA techniques within the constraints of a managed care setting.

In closing, despite the above discussed challenges and difficulties, I don't want to come across as someone who's here to declare CBA's or clinical psychology's demise. I do not believe this to be the case. A large number of my colleagues are pessimistic, but I remain optimistic about our future. I do think that clinical behavior analysts need to attend to the challenges that are facing clinical psychology and be mindful of the shifting market. However, we continue to have much to offer to the field as a whole, which has perhaps best been demonstrated by our history of effectively changing and controlling behavior in a variety of contexts. We now need to apply those skills to the context of mainstream clinical work.

## REFERENCES

- Dougher, M.J. (1993). (Ed) Special section on clinical behavior analysis: Part One. *The Behavior Analyst*, 16, 269-330.
- Dougher, M.J. (1994). (Ed) Special section on clinical behavior analysis: Part Two. *The Behavior Analyst*, 17, 287-364.
- Follette, W.C., Naugle, A.E., & Callaghan, G.M. (1996). A radical behavioral understanding of the therapeutic relationship in effecting change. *Behavior Therapy*, 27, 623-641.

- Hayes, S.C. (1995). *Acceptance and commitment therapy: A working manual for the treatment of emotional avoidance disorders*. Reno, NV: Context Press.
- Hayes, S.C. (1998a). Market-driven treatment development. *AABT the Behavior Therapist*, 21 (2), 32-33.
- Hayes, S.C. (1998b). Practice guidelines are coming - Practice guidelines are here. *AABT the Behavior Therapist*, 21 (8), 153-156.
- Hayes, S.C., Jacobson, N.S., Follette, V.M., & Dougher, M.J. (1994). (Eds.). *Acceptance and change*. Reno, NV: Context Press.
- Hayes, S.C., Wilson, K.G., Gifford, E.V., Follette, V.M., & Strosahl, K. (1996). Experimental avoidance and behavioral disorders: A functional dimensional approach to diagnosis and treatment. *Journal of Consulting and Clinical Psychology*, 64, 1152-1168.
- Kendall, P.C., & Chambless, D. L. (Eds) (1998). Special section: Empirically supported psychological therapies. *Journal of Consulting and Clinical Psychology*, 66, 3-167.
- Kohlenberg, R.J., & Tsai, M. (1991). *Functional analytic psychotherapy: Creating intense and curative therapeutic relationships*. New York: Plenum.
- National Institutes of Health. (Oct. 16-18, 1995). *Integration of behavioral and relaxation approaches into the treatment of chronic pain and insomnia*. NIH Technology Assessment Panel on Integration of Behavioral and Relaxation Approaches into the Treatment of Chronic Pain and Insomnia. Bethesda, MD.
- Paul, R.H., Marx, B.P., & Orsillo, S.M. (1999). Acceptance-based psychotherapy in the treatment of an adjudicated exhibitionist: A case example. *Behavior Therapy*, 30, 149-162.
- Raw, S.D. (1999a). Does mental health managed care violate federal antitrust laws? An interview with Joseph R. Sahid, Esq. *AABT the Behavior Therapist*, 22 (3), 53-54, 67.
- Raw, S.D. (1999b). Professional issues. *AABT the Behavior Therapist*, 22 (5), 103, 107.
- Raw, S.D. (1999c). Health insurance plans circumvent the law on mental health coverage. *AABT the Behavior Therapist*, 22 (7), 139-140.
- Skinner, B.F. (1957). *Verbal behavior*. Englewood Cliffs, NJ: Prentice-Hall.
- Smith, M.L., Glass, G.V., & Miller, T.I. (1980). *The benefits of psychotherapy*. Baltimore: The Johns Hopkins University Press.
- Strosahl, K.D., Hayes, S.C., Bergan, J., & Romano, P. (1998). Assessing the field effectiveness of Acceptance and Change Therapy: An example of the manipulated training research method. *Behavior Therapy*, 29, 35-64.
- Svartberg, M., & Stiles, T.C. (1991). Comparative effects of short-term psychodynamic psychotherapy: A meta-analysis. *Journal of Consulting and Clinical Psychology*, 59, 704-714.

**Author Note:** Address editorial correspondence to: Erik M. Augustson, Ph.D., National Cancer Institute/Division of Cancer Prevention TCRB, Executive Plaza North, RM 4087B, 6130 Executive Blvd., MSC 7337, Bethesda, MD 10892-7337. Email: [augustse@mail.nih.gov](mailto:augustse@mail.nih.gov)

**ADVANCES IN CURRICULUM-BASED MEASUREMENT: ALTERNATIVE  
RATE MEASURES FOR ASSESSING READING SKILLS IN PRE- AND  
ADVANCED READERS**

Christopher H. Skinner, Christine E. Neddenriep  
The University of Tennessee  
Kathy L. Bradley-Klug, Jenine M. Ziemann  
University of South Florida

Behavioral researchers have developed Curriculum-Based Measurement (CBM) procedures that yield a rate measure (words read correctly per minute). Investigators have shown that words correct per minute provides a valid, reliable, and sensitive measure of reading proficiency in students. Recently researchers have developed additional rate measures designed to assess pre-reading skills (Dynamic Indicators of Basic Early Literacy Skills) and skill development in advanced readers (reading comprehension rates). The current paper provides a rationale for both procedures and describes an emerging research base related to these measures. Limitations of these procedures along with directions for future research are provided.

---

Curriculum-based assessment (CBA) procedures have emerged as an alternative to traditional norm-referenced standardized achievement measures (Shapiro, 1996). Most CBA procedures use the student's current educational curricula to assess individual skills. Thus, CBA procedures assess what is taught and improve upon the lack of test-text overlap associated with norm-referenced standardized achievement tests (Bell, Lentz, & Graden, 1992; Martens, Steele, Massie, & Diskin, 1995; Shapiro & Derr, 1987). This characteristic of CBA addresses many legal, ethical, and educational concerns related to bias in assessment often associated with standardized tests (e.g., assessment results being overly dependent on extra-school learning). Because of CBA's overlap between what is taught and what is assessed, CBA procedures have clear relevance to instructional planning (Deno, 1989). For example, CBA measures can indicate skills or content students have mastered and areas that require additional work (see Skinner & Schock, 1995 for an applied example).

A number of CBA models have been developed (Shapiro, 1990). Some CBA models focus on response accuracy (e.g., Gickling & Havertape, 1981) and others include criterion-referenced measures (e.g., Blankenship, 1985; Idol, Neven, & Paolucci-Witcomb, 1986). However, the model most closely associated with behavioral psychology is known as Curriculum-Based Measurement (CBM). Although CBM procedures have been developed for assessing a variety of skills including mathematics, spelling, and writing (see Deno &

Mirkin, 1977 or Shapiro, 1996), the current paper will focus only on CBM reading procedures. After providing a brief overview, analysis, and research summary of traditional CBM reading measures, recent advances in CBM research will be described along with directions for future research.

#### **Words Correct Per Minute**

CBM reading procedures yield measures of oral reading fluency. During CBM assessment procedures, students read aloud for one minute from a passage within their reading text or series while the examiner marks errors (e.g., mispronunciations, substitutions, omissions, skipped lines). The examiner calculates two rate measures: words read correctly per minute and errors per minute (Deno & Mirkin, 1977; Shapiro, 1996). Typically, words correct per minute serves as the primary measure for making educational decisions (e.g., placement in the curricula, evaluation of intervention effectiveness).

Researchers have shown that words correct per minute is a sensitive, reliable, and valid measure that can be used to assess students' general reading proficiency and to evaluate progress within students' reading curriculum (Deno, Mirkin, & Chiang, 1982; Fuchs & Deno, 1992; Fuchs & Fuchs, 1992; Fuchs, Fuchs, & Maxwell, 1988; Madelaine & Wheldall, 1999; Marston, 1989; Shinn, Good, Knutson, Tilly, & Collins, 1992).

Several characteristics of CBM make words correct per minute ideal for assessing reading skills (Shapiro, 1996; Shinn, 1995). CBM procedures are brief and inexpensive. Multiple measures can be constructed from a student's curriculum. Because students are assessed using probes (i.e., brief passages) taken from their curricula, issues related to test-text overlap are mitigated. These characteristics of CBM procedures allow for repeated assessments over brief periods of time, permitting educators to monitor a student's progress, evaluate the effects of academic interventions, and compare the effects of interventions (Daly & Martens, 1994; Forness, Kavale, Blum, & Lloyd, 1997; Shinn, 1995; Skinner, Cooper, & Cole, 1997). While norm-referenced standardized achievement tests also allow researchers to make similar decisions, they have minimal value to practitioners because they cannot be given often enough to allow educators to decide whether they should a) adjust a student's curriculum (e.g., place in higher level reading material), b) adjust a student's placement (e.g., move a student to a more advanced reading group), or c) alter current instructional procedures (Shapiro, 1996).

Each of these applied advantages of CBM procedures is dependent upon their sensitivity. Frequent assessment is of little value unless the assessment procedures are sensitive enough to detect small changes in an individual student's skills over brief periods of time. The characteristic of CBM that accounts for sensitivity is that CBM measures *rates* (Shinn, 1995). Perhaps, because CBM data are rate based, CBM has received much attention and support from educators with a behavioral background. After all, the whole of B. F. Skinner's basic operant research was based on patterns of behavior that emerged when rates of behavior were measured (e.g., Ferster & Skinner, 1957).

### Research Evaluating Words Correct Per Minute

A strong research base exists supporting CBM procedures. Researchers have investigated the criterion-related validity of words correct per minute (Deno et al., 1982; Fuchs et al., 1988; Jenkins & Jewel, 1993; Marston, 1989). These

studies have correlated oral reading fluency (words correct per minute) as well as other alternative measures (e.g., question answering, recall, cloze, maze) concurrently with established norm-referenced tests of reading. For example, Deno et al. (1982) administered two oral reading tasks (i.e., reading words in isolation and reading passages), one cloze comprehension task, and selected subtests from two norm-referenced achievement tests (i.e., Stanford Diagnostic Reading Test [SDRT] and the Peabody Individual Achievement Test [PIAT]) to a total of 66 general and special education students in first through sixth grades. Correlations between oral reading and both the literal and inferential comprehension subtests of the SDRT were .78 and .80, while the correlations between cloze comprehension and the same subtests of the SDRT were lower (.67 and .71). Based on these correlations, Deno et al. concluded that reading aloud could be used to assess both decoding and comprehension proficiency in students. Similarly, Fuchs et al. (1988) administered the oral passage reading test along with three informal measures of reading comprehension (i.e., question answering, written recall, and written cloze) and selected subtests from the Stanford Achievement Test (SAT) to a total of 70 mild to moderately impaired students in fourth through eighth grades. The oral reading task was found to correlate more strongly with standardized test scores than any of the written informal measures of reading comprehension (.91 with the SAT Reading Comprehension Subtest).

This relationship between oral reading fluency and reading comprehension has been found to remain strong across differing types of curricula (e.g., traditional basal, literature-based, and authentic reading materials) and difficulty levels (Fuchs & Deno, 1992; Hintze, Shapiro, Conte, & Basile, 1997). In addition, generalizability theory has been used to evaluate the reliability and validity of oral reading fluency in making treatment decisions (i.e., to assess its dependability across varying sources of error such as curriculum and time). These studies investigating generalizability theory support the use of oral reading fluency in progress monitoring when making comparisons

both within and between individuals (Hintze, Owen, Shapiro, & Daly, 2000).

Other studies provide support for the construct validity of CBM measures. Kranzler, Brownell, and Miller (1998) used multiple regression analyses to determine whether the relationship between oral reading fluency and comprehension could be explained by differences in cognitive ability. Results of the multiple regression analyses validated the construct validity of oral reading fluency. Despite the inclusion of measures of general cognitive ability, processing speed, and efficiency in the multiple regression analyses, the contribution of oral reading fluency to the prediction of reading comprehension was found to be significant (Kranzler et al.). Thus, this study provided additional support for the relationship between oral reading fluency and comprehension.

A large database exists supporting the construct and criterion validity of CBM procedures. However, two related concerns may need to be addressed. Although many correlations with criterion measures are strong, Mehrens and Clarizio (1993) indicate that these results may be inflated because researchers have computed correlations across multiple grade levels, increasing within group variability. To address this limitation, Jenkins and Jewell (1993) administered an oral reading passage task, a maze task, and selected subtests from two norm-referenced achievement tests (Gates-MacGinitie Reading Tests and the Metropolitan Achievement Test [MAT]) to 335 students from second to sixth grade. The correlation between oral reading fluency and the Gates-MacGinitie total reading score was found to be comparable with previous studies (.86). When correlations were computed with respect to grade level, however, results were found to be variable. In grades two through four, correlations between oral reading fluency and both the total reading score and comprehension remained strong (Gates-MacGinitie, range .82 to .88). At grades fifth and sixth, however, the relationship was found to decrease (range .63 to .73).

Jenkins and Jewell (1993) suggest that this decline in the strength of the relationship

can be explained by the progressive change in the developmental tasks of reading. According to Chall's (1983) developmental theory, students' reading skills progress from an ability to decode words and the development of fluency to the ability to comprehend and integrate material. While achievement tests progressively assess skills related to comprehension and general knowledge at the intermediate grades, words correct per minute may not reflect this same developmental growth in reading skill.

Hintze and Shapiro (1997) also found differences in growth by age when using words correct per minute. Their study showed that growth in words correct per minute was linear until grade five at which a drop in growth was observed. The decreasing trend in correlations between words correct per minute and reading comprehension as students' reading skills improve remains problematic. This ceiling effect may be caused by development changes in reading skills (e.g., Chall, 1983) or mere limits on individuals' ability to read aloud at more rapid rates. Regardless, these data suggest that CBM procedures may not be as useful (e.g., may be a less sensitive measure) for making educational decisions regarding students reading above the fourth-grade level.

A second, related concern is what oral reading fluency *appears* to measure and educators' acceptance of this measure of reading proficiency. These are issues of face validity and social validity (Fuchs & Fuchs, 1992). Teachers' beliefs about a measure influence their willingness to use the measure with their students (Gersten, Woodward, & Morvant, 1992). Many educators doubt that oral reading fluency actually measures comprehension (Fuchs & Fuchs). On the surface, oral reading fluency would appear to measure the ability to decode words quickly and accurately (Potter & Wamre, 1990; Shinn et al., 1992). While the ability to decode words may be a necessary prerequisite for comprehension, efficient decoding alone may not be sufficient for a student to comprehend what they read (Perfetti & Hogaboam, 1975).

Shinn et al. (1992) completed a confirmatory factor analysis to determine the

theoretical contribution of oral reading fluency to the process of reading. Shinn et al. administered a total of eight measures to 114 third-grade students and 124 fifth-grade students. Two of the measures assessed decoding skills; two measures assessed oral reading fluency; and, four measures assessed reading comprehension. For the third-grade students, a single factor model, labeled "reading competence" provided the most parsimonious explanation of the reading process, with all measures contributing significantly to this single factor. The measures of oral reading fluency (words correct per minute) correlated most strongly with this single-factor model (.88 and .90). For the fifth-grade students, however, a two-factor model was found to best depict the reading process, decoding and comprehension. While decoding and comprehension were found to be best described as separate constructs these measures were highly correlated (.83). As well, the oral reading fluency measures were found to load robustly on the reading decoding factor. Thus, the relationship between oral reading fluency and reading comprehension at the fifth grade level may best be described as a strong, but indirect relationship.

#### **Emerging Research on Rates of Comprehension: An Upward Extension of CBM**

A large number of studies suggest that words correct per minute is a useful measure of general reading proficiency for students reading in grade levels two to four. However, as students' reading skills develop beyond grade four, CBM may suffer from three limitations. First, CBM data reflects less growth in words correct per minute as students' reading skills improve. Thus, CBM may lose some of its sensitivity to growth for more skilled (e.g., older) readers. Second, researchers have shown that the correlation between CBM measures and reading comprehension declines as students' reading skills improve. Finally, the target behavior (words correct per minute) may lack both face and educational validity for more skilled readers (Chall, 1983; Fuchs & Fuchs, 1992; Potter & Wamre, 1990; Skinner, 1998).

#### **Reading Comprehension Rates**

Given the limitations of measuring words correct per minute, another metric may be useful for assessing more skilled readers (e.g., students beyond grade four). One alternative that has been recently investigated is to more *directly* measure reading comprehension rates. While many measures have been developed to measure comprehension, few have been developed to address *rates* of comprehension.

Skinner (1998) provides a rationale for why rates of comprehension are educationally valid measures of reading skills, particularly in students beyond the fourth grade. Skinner's rationale is based on an assumption regarding the function of reading and basic behavioral research on choice behavior. The primary assumption is that individuals read for comprehension. In behavioral terms, the function of reading would be comprehension. Comprehension is often reinforced. For example, students who read and comprehend well may receive a higher test grade or perform better on laboratory assignments because they have read and comprehended material. Additionally, other reinforcement associated with reading for pleasure (e.g., understanding humor or foreshadowing) requires readers to comprehend. If reinforcement is delivered contingent upon comprehension, then those who comprehend well are likely to benefit more from their reading behaviors (e.g., enjoy a novel better because they comprehend the subtle foreshadowing).

While comprehension is critical, comprehension alone is not sufficient (Skinner, 1998). For example, suppose two eighth-grade students are given a history passage and told to read it. Now assume that both students read the passage and comprehend 80% of it. Therefore, both get 80% on an exam. Now, also assume that for one student the passage took 10 minutes to read and the other student required 30 minutes to read the passage. The rate of reinforcement for the student who read the passage in 10 minutes, all things being equal, is 3 times higher than the student who read the passage in 30 minutes. Also, the more rapid reader has additional time available to engage in other behavior. For the eighth-grade student, this may

include studying for a science test or attending swim practice. For the adult businessperson, this may include making several more sales calls, increasing their productivity.

### Measuring Reading Comprehension Rate

Converting oral reading accuracy (e.g., percent words read correctly) to a rate measure (e.g., words correct per minute) enhances sensitivity of oral reading measures. However, as students' reading skills develop, they are rarely reinforced for rapid and accurate *oral* reading. Exceptions include those who read the news, make commercials for the radio, or give speeches; they may be reinforced for rapid accurate aloud reading (Freeland, Skinner, Jackson, McDaniel, & Smith, 2000). In most instances, reinforcement for reading is delivered contingent upon comprehension. Therefore, researchers have begun to develop rate measures of reading comprehension. As with words correct per minute, converting comprehension data to rate measures should enhance the sensitivity of these measures. Additionally, such a measure provides a more direct and educationally valid measure of reading comprehension.

Several studies have been conducted where rates of reading comprehension were used to assess the effects of reading interventions (Freeland et al., 2000; Jackson, Freeland, & Skinner, 2000; McDaniel et al., 2001; Neddenriep, Skinner, Abramson, & Wallace, 2002). In each of these studies, one of two similar curricula was used: *Timed Readings* (Spargo, 1989a) and *Timed Readings in Literature* (Spargo, 1989b). Each of these curricula contains 10 books, beginning at the fourth-grade reading level. Each book contains 50 passages that become progressively more difficult. Each passage contains 400 words and 10 multiple choice comprehension questions: five inferential questions and five factual questions.

With traditional CBM measures, words correct per minute is in the numerator and time is in the denominator. In order to measure rates of comprehension, researchers merely replace words correct in the numerator with the percent

of comprehension questions answered correctly, while keeping time in the denominator. Using Spargo's (1989a,b) 400-word passages, a corollary measure of reading comprehension rates would be the proportion of the passage comprehended for each minute spent reading. Percent comprehension per minute (%C/M) would be calculated using the following formula:

$$\%C/M = (\% \text{ comprehension questions correct}) / (\text{number of seconds spent reading})$$

Thus, for a student who read a 400-word passage in 6 minutes and answered 80% of the comprehension questions correctly, his %C/M would equal 13.3%. This means that for each minute spent reading the student understood 13.3% of the passage. Similarly, for a student who read the passage in 3 minutes and answered 80% of the comprehension questions correctly, %C/M would equal 26.7%. This means that for each minute spent reading, the student understood 26.7% of the passage.

### Reading Comprehension Rates: An Emerging Research Base

To date, several studies have used rate of comprehension to evaluate the effects of reading interventions. Freeland et al. (2000) used a multi-element design to evaluate the effects of a repeated reading intervention on *silent* reading comprehension rates in three secondary students with learning disabilities. This study showed that the measure was sensitive and stable enough to detect differences across the two conditions (i.e., repeated reading and control conditions). Additionally, results indicated that repeated readings enhanced student's factual comprehension rates, but not inferential comprehension levels or rates. In a similar study, Jackson et al. (2000) found that this measure also was sensitive and stable enough to detect increases in reading comprehension rates in secondary students associated with a listening-while-reading intervention.

While these studies provide some initial support for the silent reading comprehension rate measure, another study indicated clear limitations associated with this measure.



McDaniel et al. (2001) used adapted alternating treatments designs in an attempt to compare the effects of listening-while-reading and repeated reading on secondary students' silent reading comprehension rates. After being exposed to either condition the previous day, the students were instructed to re-read the passage silently, report when they had finished, and then answer the comprehension questions. Unfortunately, data on time required to re-read the passages suggested that students frequently did not silently re-read the passage. Specifically, the time they spent silently reading was so brief that they could not have re-read the passage silently.

The McDaniel et al. (2001) study indicates a serious limitation of measuring rates of silent reading comprehension: educators and researchers cannot be certain that the students actually read the entire passage silently. One solution to this problem is to require students to read the passages orally. In a series of studies evaluating classwide peer tutoring components, Neddenriep et al. (2002) evaluated interventions by having sixth-grade students read equivalent passages orally and measured a) words correct per minute (WC/M); b) percent comprehension questions correct; and c) rates of oral, as opposed to silent reading comprehension. Correlational data showed that comprehension rates correlated more strongly with words correct per minute than comprehension levels (i.e., percent correct). Specifically, the correlation between WC/M and comprehension levels (i.e., percent questions correct) was .57, while the correlation between WC/M and comprehension rate was .87.

Taken together, these studies suggest that when measuring rates of comprehension students should be required to read aloud, as opposed to silently. However, one limitation associated with this procedure is that the measure may lose some face validity because as students become more skilled readers, they read silently for comprehension. Clearly more research is needed to investigate the validity, reliability, and sensitivity of reading comprehension rates, when reading both orally and silently.

CBA procedures are said to be more direct than traditional standardized assessment procedures because the material comes directly from the students' curricula. Perhaps the most serious limitation associated with measuring rates of comprehension is that such measures require specialized materials. Past researchers used Spargo's (1989a,b) *Timed Readings* curricula to investigate rates of comprehension because a) each passage contained 400 words, b) passages were constructed to become progressively more difficult, and c) each passage contained 10 multiple choice comprehension questions. These characteristics allow researchers to measure comprehension rates across same length passages. Commonly used reading series or literature curricula contain passages of varying length, which do not lend themselves to repeated measures.

A related problem concerns passage context and comprehension. When measuring words correct per minute a probe can be constructed by selecting a portion of a passage (e.g., 200 words from a story). However, reading mere portions of a passage may not allow the reader to fully comprehend material, as this portion has been taken out of context. Because passages often used for measuring words correct per minute do not have a clear beginning, middle, and end, they do not lend themselves to assessing higher-level comprehension skills (e.g., selecting main ideas, integrating material from beginning and end of passage). The Spargo series (1989a,b) were used to assess rates of comprehension because readers could complete each 400-word passage quickly and evaluators could assess higher-level comprehension (e.g., inferential questions requiring comprehension of an entire discrete passage).

CBM reading procedures were originally designed for use with traditional basal reading series. However, a number of changes in reading instruction have occurred over the years and as a result, the curricula used for instruction have changed. Literature-based curriculum and authentic literature have replaced traditional basal texts in many reading programs. The question then becomes whether CBM reading probes must truly be curriculum-based.

Recent research suggests that the issue of curricula overlap may be less important than previously thought. Researchers have used CBM probes taken from alternative reading curricula (i.e., not the curricula in which students are being instructed) and evaluated their usefulness for assessing reading progress (Bradley-Klug, Shapiro, Lutz, & DuPaul, 1998; Hintze & Shapiro, 1997; Powell-Smith & Bradley-Klug, 2001). Results showed that these CBM probes yielded similar data when assessing student progress (i.e., words correct per minute over 8-10 weeks were similar regardless of curricula overlap). These studies suggest that words correct per minute may be a valid, reliable, and sensitive measure of students' reading skill development, even when they are reading from an alternative series. If future researchers find similar results with rates of comprehension measures, then it may be possible to develop specific material used to assess reading comprehension rates across students [e.g., use Spargo's (1989a,b) series to assess comprehension rates in students working from other curricula]. Additionally, as students' reading skills progress, overlap may become less of an issue as reading instruction gives way to reading appreciation and comprehension of content, as opposed to accurate and rapid reading (Chall, 1983). Future research needs to be conducted to determine the impact of overlap on rates of reading comprehension across reading skill levels.

Additional research on comprehension questions also is needed. The Spargo (1989a,b) series includes 10 multiple choice comprehension questions. In order to use rates of comprehension to measure reading skill improvement, comprehension questions must be equated across passages or become progressively more difficult as the curricula advances. Additionally, students are likely to have different background knowledge with respect to the content or topic of the passages. Taking the median of three measures across three passages has been used to address these issues when measuring words correct per minute (Deno & Mirkin, 1977). Future researchers should determine if similar procedures would improve the reliability of comprehension rate measures.

Although there are many limitations associated with measuring comprehension rates, these measures may provide a more sensitive and more educationally valid measure of reading skills in more skilled readers than words correct per minute. Research on this measure has just begun. Clearly, more research is needed before such a measure can be employed for educational decision-making.

#### **Emerging Research on Early Literacy Rate Measures: A Downward Extension of CBM**

In addition to developing alternative rate measures for assessing reading in more skilled readers, researchers have begun to investigate and develop alternative measures for less skilled readers (students reading at the first or pre-first grade level). The problem with using words correct per minute with less skilled readers is clear: students have not yet developed their reading skills to the point where they can read passages orally (Good, Simmons, & Smith, 1998). Like CBM, these procedures do more than assess accuracy: they assess rates.

Developing sensitive, reliable, and valid measures of pre-reading skills that can be administered frequently may be critical to preventing and remedying reading problems. Longitudinal studies of reading acquisition have repeatedly found that reading performance in early years is highly correlated with reading skills in later years (Good et al., 1998). In particular, students with initially high pre-reading skills (e.g., phonological awareness, alphabetic understanding) are able to acquire oral reading fluency more quickly, while those students with initially low early literacy skills acquire reading fluency at a slower rate (Stanovich, 1986). As a result, students who emerge in early years as less-skilled readers are likely to continue to have lower reading skills throughout their education. Stanovich referred to this phenomenon as the "Matthew Effect," in which the reading-rich get richer and the reading-poor get poorer (see also Hart & Risley, 1995). In order to catch up with their peers, students with low initial early literacy skills would have to improve their reading fluency (words correct per minute) at a rate 5 times faster than the average student (Good et al., 1998). One solution is to intervene early and

ensure that students have adequate pre-reading skills, so that they are sufficiently prepared for interaction with connected text as they progress in reading skill development.

Although CBM has been successfully used to identify and design interventions for students with established reading difficulties, this metric is less sensitive to the development of skills essential to early literacy. At the beginning of first grade, students with high and low early literacy skills are indistinguishable using reading CBM because neither group has measurable oral reading fluency (Good et al., 1998). Direct measures of reading, (e.g., words correct per minute), therefore, are insufficient for identifying students at risk for reading difficulties because they cannot identify problems until they are well-established and resistant to change.

#### **Dynamic Indicators of Basic Early Literacy Skills (DIBELS)**

In order to determine which children are at risk for developing poor reading skills, a set of measures has been created that apply a prevention-oriented, problem-solving approach to identification of reading difficulties. Dynamic Indicators of Basic Early Literacy Skills (DIBELS) were developed by researchers at the University of Oregon as a downward extension of CBM and were designed to be brief, easily administered, and technically adequate measures of younger children's early literacy skills (Good & Kaminski, 1996; Kaminski & Good, 1998). DIBELS assessment procedures can be used to identify which students have insufficient early literacy skills, formatively evaluate student progress towards literacy development, and determine when interventions have successfully reduced risk of reading failure by improving early literacy skills.

DIBELS are *dynamic* measures that are sensitive to improvements in students' early literacy over time, as well as to changes in performance that result from reading interventions. To be capable of measuring student change, dynamic measures must be easy to administer, capable of frequent, repeated administration, and time/cost effective (Kaminski & Good, 1996). As with CBM,

DIBELS are intended to be *indicators*, or basic vital signs of students' educational growth in the critical areas of early literacy. DIBELS measures are to be used as a fast, efficient way to alert teachers and parents to problems in students' pre-reading skills (Kaminski & Good, 1996).

Although DIBELS parallel CBM in their rationale and basic procedure, they diverge in the areas of reading skills assessed. Rather than directly measuring reading fluency, DIBELS are designed to focus on *basic early literacy skills* that are prerequisite foundations for reading success. The validity and utility of such early literacy measures are grounded in two key characteristics: predictive validity for future reading performance and a functional relationship with reading acquisition (Good et al., 1998; Good, Kaminski, Simmons, & Kame'enui, 2001).

DIBELS assess three of the skill areas among the "big ideas in early reading:" a) *phonological awareness*, or the ability to hear and manipulate the sound structure of language; b) *alphabetic principle*, or the ability to coordinate print with speech, recode strings of letters into sounds, and blend sounds into words; and c) *fluency with text*, or the ability to quickly, accurately, and automatically decode connected words. These "big ideas" can be conceptualized as a series of successive steps that build upon each other to yield reading success. DIBELS measures do not comprehensively measure all aspects of phonological awareness, alphabetic principle, and fluency with text, but rather tap into one or more specific skills pertaining to each critical area to provide an index of student performance. Early competencies in these particular skill areas have been found to differentiate successful from less successful readers and can be improved through remediation and instruction (Good et al., 2001). Each of the DIBELS measures is individually administered to the child and has been standardized to establish technical adequacy. A brief description of the purpose and procedures for each DIBELS measure is provided below.

*Initial Sounds Fluency (ISF)*. DIBELS Initial Sound Fluency (ISF; formerly known as

DIBELS Onset Recognition Fluency or OnRF) assesses a child's ability to detect and produce the initial sound in an orally-presented word (Good et al., 2001). To administer ISF to a child, the examiner presents four pictures to the child, names each of them, and asks the child to point to or name the picture that begins with the sound orally produced by the examiner (e.g., "This is sink, cat, gloves, and hat. Which picture begins with /s/?"). The child also is asked to provide the initial sound for several of the stimulus pictures (e.g., "What sound does *hat* begin with?"). ISF takes approximately three minutes to administer and is scored by calculating the time a child takes to identify the correct picture or produce the correct sound, and determining the number correct in a minute. Over 20 alternate forms of ISF are available for repeated testing and progress monitoring.

*Letter Naming Fluency (LNF).* DIBELS Letter Naming Fluency (LNF) functions as a general indicator of student risk. A student is presented a page of randomly arranged upper- and lower-case letters and is told to name as many as he/she can. If the student does not know a letter, the examiner will provide it. The student has one minute to pronounce as many letter names as possible, resulting in a score of letters correctly named per minute. Student performance is gauged in comparison to peers within their school district, and a score below the 20<sup>th</sup> percentile on district norms suggests that the student may be at risk for difficulty in meeting early literacy benchmark goals. Students who score between the 20<sup>th</sup> and 40<sup>th</sup> percentile on LNF are considered to be at some risk for reading difficulty, and those who score above the 40<sup>th</sup> percentile in their district are considered to be at low risk.

*Phoneme Segmentation Fluency (PSF).* DIBELS Phoneme Segmentation Fluency (PSF) assesses students' phonemic awareness, particularly their ability to fluently segment three- and four-phoneme words into their individual phonemes. Performance on the PSF measure has been found to be a good predictor of future achievement in reading (Kaminski & Good, 1996). The procedure for PSF consists of the examiner orally presenting three- and four-phoneme words to the student, who is required

to verbally produce the individual phoneme of each word (e.g., examiner says "cat," student says "/c/ /a/ /t/"). The student's final score reflects the number of correct phonemes produced in one minute. Total administration time is approximately two minutes, and over 20 alternate forms of PSF are available for progress monitoring.

*Nonsense Word Fluency (NWF).*

DIBELS Nonsense Word Fluency (NWS) is a measure of students' understanding of the alphabetic principle, or the concept that words are composed of letters that represent sounds. In particular, NWF assesses letter-sound correspondence and sound blending skills by requiring students to pronounce a series of random vowel-consonant or consonant-vowel-consonant combinations (e.g., *ov*, *sig*, *rav*). Students may either pronounce each sound of the word (e.g., /s/ /i/ /g/) or may read the whole nonsense word (e.g., "sig"). For this example, the student would receive 3 points, as each correctly pronounced sound is counted as a point. The student has one minute to produce as many letter-sounds as possible, resulting in a letter-sounds per minute score. Because NWF assesses fluency of letter-sound production, students receive a higher score when they phonologically recode the word by reading it whole, because they are able to produce more letter-sounds by the minute time limit. Students who produce the sounds in isolation demonstrate less fluency in letter-sound production and thus receive a lower score.

### Emerging Research Base for DIBELS

Despite their relative infancy, DIBELS measures have a fair amount of research to support their use. Technical adequacy studies have established the reliability and validity of the measures (Kaminski & Good, 1996). Alternate-form reliability for the ISF, LNF, PSF, and NWF measures ranges from .72-.88. These levels of reliability, while low by traditional statistical standards, are considerably high for one-minute, repeatable measures. Concurrent-/criterion-related validity has been assessed by comparing students' performance on DIBELS to performance on the Woodcock-Johnson Psychoeducational Battery Readiness Cluster

score, and ranges from .36-.59. Predictive validity was established in comparison to scores on the Woodcock-Johnson Psychoeducational Battery Total Reading Score (.36-.68) and to fluency rates from CBM in first grade (.36-.82).

DIBELS have been found to be adequately sensitive to students' changes and growth over time. A study by Kaminski and Good (1996) found that first-grade children scored significantly higher than kindergarten children on all DIBELS measures, suggesting that their performance on these measures changes considerably over the course of an academic year. Slope data of students' performance on repeated administrations of PSF also showed considerable growth over a 9-week period, which indicate sensitivity to student progress in pre-literacy skills. These findings support the use of DIBELS in monitoring progress toward literacy and evaluating intervention efficacy.

Case study research also supports the use of DIBELS for instructional decision-making within a problem solving-model. Good and Kaminski (1996) describe a kindergarten student who was identified as having below-average phonological awareness skills through DIBELS ISF and PSF. These measures continued to be administered as intervention options for the student were explored, and DIBELS data showed his improvement in ISF and PSF as a result of intervention. At the end of the academic year, DIBELS data were used to evaluate intervention efficacy by comparing the target student's performance to the median of his kindergarten. Although this study provides an instructive illustration of DIBELS' utility in educational decision-making, there are few other empirical demonstrations of DIBELS to date. Group and single-subject studies of these measures offer a promising line of research in years to come, and will further elucidate the instructional benefits they offer.

Individualized, case-by-case identification of student problems, progress monitoring, and evaluations of interventions are key to using DIBELS in a problem-solving context. The combined use of CBM and DIBELS allows educational decision-making to

be guided by an Outcomes-Driven Model which is prevention-oriented and designed to remedy reading difficulties before they become resistant to intervention (Good et al., 2001). Brief, reliable, and valid measures of empirically-validated "big ideas" in reading empower educators to engage in a series of decision-making steps aimed at individual student success. Ultimately, the goal of the Outcome-Driven Model, as well as its accompanying assessment instruments, is to match student needs to appropriate instructional supports *before* a pattern of reading failure becomes established.

### Summary and Conclusions

Traditional measures of assessing academic achievement focus on accuracy of academic responding. Behavioral researchers have developed Curriculum-Based Measurement procedures, which measure *rates* of accurate overt responding (e.g., words read correctly per minute). A large research base has shown that words correct per minute is a valid and reliable measure of general reading proficiency. Additionally, CBM procedures lend themselves to frequent assessment, and words correct per minute is sensitive enough to allow educators to evaluate instructional decisions in order to maximize reading skill development of students in grades two through four (Forness et al., 1997).

Although researchers have repeatedly shown that words read correctly per minute is a useful measure of reading fluency (i.e., rapid and accurate reading), this measure may be less sensitive and educationally valid when assessing more skilled readers; and, words read correctly per minute cannot be used to assess pre-readers. In the current paper, two promising measures of reading skill fluency are described: reading comprehension rates and DIBELS. Both measures assess rates. However, the target behavior is altered to better address students' current level of reading skill development (Chall, 1983; Good et al., 1998; Skinner, 1998).

Reading comprehension rates and DIBELS are both based on behavioral assessment principles (Hartmann, Roper, & Bradford, 1979) in that they directly measure

target behaviors and employ rate measures. Also, both are new measures that require additional research. Given the history of behavioral researchers, it is likely that both measures will generate future research designed to evaluate and enhance these emerging assessment procedures.

## REFERENCES

- Bell, P. F., Lentz, F. E., & Graden, J. L. (1992). Effects of curriculum-test overlap on standardized tests scores: Identifying systematic confounds in educational decision making. *School Psychology Review, 21*, 244-255.
- Blakenship, C. S. (1985). Using curriculum-based assessment to make instructional management decisions. *Exceptional Children, 42*, 233-238.
- Bradley-Klug, K. L., Shapiro, E. S., Lutz, J. G., & DuPaul, G. J. (1998). Evaluation of oral reading rate as a curriculum-based measure within literature-based curriculum. *Journal of School Psychology, 36*, 183-197.
- Chall, J. S. (1983). *Stages of reading development*. New York: McGraw-Hill.
- Daly, E. J., & Martens, B. K. (1994). A comparison of three interventions for increasing oral reading performance: Application of the instructional hierarchy. *Journal of Applied Behavior Analysis, 27*, 459-469.
- Deno, S. L. (1989). Curriculum-based measurement and special education services: A fundamental and direct relationship. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 1-17). New York: Guilford.
- Deno, S. L., & Mirkin, P. K. (1977). Data-based problem modification: A manual. Reston, VA: Council for Exceptional Children.
- Deno, S. L., Mirkin, P. K., & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children, 49*, 36-45.
- Ferster, C. B., & Skinner, B. F. (1957). *Schedules of reinforcement*. New York: Appleton-Century-Crofts.
- Fomess, S. R., Kavale, K. A., Blum, I. M., & Lloyd, J. W. (1997). Mega-analysis of meta-analyses: What works in special education and related services. *Teaching Exceptional Children, 29*(6), 4-9.
- Freeland, J. T., Skinner, C. H., Jackson, B., McDaniel, C. E., & Smith, S. (2000). Measuring and increasing silent reading comprehension rates via repeated readings. *Psychology in the Schools, 37*, 415-429.
- Fuchs, L. S., & Deno, S. L. (1992). Effects of curriculum within curriculum-based measurement. *Exceptional Children, 58*, 232-243.
- Fuchs, L. S., & Fuchs, D. (1992). Identifying a measure for monitoring student reading progress. *School Psychology Review, 21*, 45-58.
- Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education, 9*(2), 20-28.
- Gersten, R. M., Woodward, J., & Morvant, M. (1992). Refining the working knowledge of experienced teachers. *Educational Leadership, 49*, 34-39.
- Gickling, E. E., & Havertape, J. (1981). *Curriculum-based assessment*. Minneapolis, MN: National School Psychology Inservice Training Network.
- Good, R. H., III, & Kaminski, R. A. (1996). Assessment for instructional decisions: Toward a proactive/preventative model of decision making for early literacy skills. *School Psychology Quarterly, 11*, 326-336.
- Good, R. H., III, Kaminski, R. A., Simmons, D. C., & Kame'enui, E. J. (2001). Using dynamic indicators of basic early literacy skills (DIBELS) in an outcomes-driven model: Step to reading outcomes. *Oregon School Study Council Bulletin, 44*(1). (ERIC Document Reproduction Service No. ED453526).
- Good, R. H., III, Simmons, D. C., & Smith, S. B. (1998). Effective academic interventions in the United States: Evaluating and enhancing the acquisition of early reading skills. *School Psychology Review, 27*, 45-56.
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Paul E. Brookes.
- Hartmann, D. P., Roper, B. L., & Bradford, C. C. (1979). Some relationships between behavioral and traditional assessment. *Journal of Behavioral Assessment, 1*, 3-21.
- Hintze, J. M., Owen, S. V., Shapiro, E. S., & Daly, E. J., III. (2000). Generalizability of oral reading fluency measures: Application of G theory to curriculum-based measurement. *School Psychology Quarterly, 15*, 52-68.
- Hintze, J. M., & Shapiro, E. S. (1997). Curriculum-based measurement and literature-based reading: Is curriculum-based measurement meeting the needs of changing reading curricula? *Journal of School Psychology, 35*, 351-375.
- Hintze, J. M., Shapiro, E. S., Conte, K. L., & Basile, I. M. (1997). Oral reading fluency and authentic reading material: Criterion validity of the technical features of CBM survey-level assessment. *School Psychology Review, 26*, 535-553.
- Idol, L., Neven, A., & Paolucci-Witcomb, P. (1986). *Models of curriculum-based assessment*. Rockville, MD: Aspen.
- Jackson, B. J., Freeland, J. T., & Skinner, C. H. (2000, Nov.). *Using reading previewing to improve reading comprehension rates for secondary student with reading deficits*. Paper presented at the annual meeting of the Mid-South Educational Research Association, Bowling Green, KY.
- Jenkins, J. R., & Jewell, M. (1993). Examining the validity of two measures for formative teaching: Reading aloud and maze. *Exceptional Children, 59*, 421-432.
- Kaminski, R. A., & Good, R. H., III (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review, 25*, 215-227.
- Kaminski, R. A., & Good, R. H., III (1998). Assessing early literacy skills in a problem-solving model: Dynamic indicators of basic early literacy skills. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 113-142). New York: Guilford Press.
- Kranzler, J. H., Brownell, M. R., & Miller, M. D. (1998). The construct validity of curriculum-based measurement of

- reading: An empirical test of a plausible rival hypothesis. *Journal of School Psychology, 36*, 399-415.
- Madelaine, A., & Wheldall, K. (1999). Curriculum-based measurement of reading: A critical review. *International Journal of Disability, Development and Education, 46*, 71-85.
- Martens, B. K., Steele, E. S., Massie, D. R., & Diskin, M. J. (1995). Curriculum bias in standardized tests of reading decoding. *Journal of School Psychology, 33*, 287-296.
- Marston, D. B. (1989). Curriculum-based measurement approach to assessing academic performance: What it is and why do it. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 18-78). New York: Guilford.
- McDaniel, C. E., Watson, T. S., Freeland, J. T., Smith, S. L., Jackson, B., & Skinner, C. H. (2001, May). *Comparing silent repeated reading and teacher previewing using silent reading comprehension rate*. Paper presented at the annual Convention of the Association for Behavior Analysis, New Orleans, LA.
- Mehrens, W. A., & Clarizio, H. F. (1993). Curriculum-based measurement: Conceptual and psychometric considerations. *Psychology in the Schools, 30*, 241-254.
- Neddeniep, C. E., Skinner, C. H., Abramson, E. M., & Wallace, M. A. (2002, May). *The effects of classwide peer tutoring on oral reading fluency and silent reading comprehension*. Poster presented at the Annual Convention of the Association for Behavior Analysis; Toronto, Canada.
- Perfetti, C. A., & Hogaboam, T. (1975). Relationship between single word decoding and reading comprehension skill. *Journal of Educational Psychology, 67*, 461-469.
- Potter, M. L., & Wamre, H. M. (1990). Curriculum-based measurement and developmental reading models: Opportunities for cross-validation. *Exceptional Children, 57*, 16-25.
- Powell-Smith, K. A., & Bradley-Klug, K. L. (2001). Another look at the "C" in CBM: Does it really matter if curriculum-based measurement reading probes are curriculum-based? *Psychology in the Schools, 38*, 299-312.
- Shapiro, E. S. (1990). An integrated model for curriculum-based assessment. *School Psychology Review, 19*, 331-349.
- Shapiro, E. S. (1996). *Academic skills problems: Direct assessment and intervention* (2nd ed.). New York: Guilford Press.
- Shapiro, E. S., & Derr, T. F. (1987). An examination of overlap between reading curricula and standardized achievement tests. *Journal of Special Education, 21*, 59-67.
- Shinn, M. R. (1995). Curriculum-based measurement and its use in a problem solving model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology-III* (pp. 547-567). Washington DC: National Association of School Psychologists.
- Shinn, M. R., Good, R. H., Knutson, N., Tilly, W. D., & Collins, V. L. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review, 21*, 459-479.
- Skinner, C. H. (1998). Preventing academic skills deficits. In T. S. Watson & F. Gresham (Eds.), *Handbook of child behavior therapy* (pp. 61-83). New York: Plenum.
- Skinner, C. H., Cooper, L., & Cole, C. L. (1997). An examination of rapid and slow rate listening previewing interventions on reading performance. *Journal of Applied Behavior Analysis, 30*, 331-333.
- Skinner, C. H., & Schock, H. H. (1995). Best practices in mathematics assessment. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology-III* (pp. 547-567). Washington DC: National Association of School Psychologists.
- Spargo, E. (1989a). *Time readings*. Providence, RI: Jamestown Publishers.
- Spargo, E. (1989b). *Time readings in literature*. Providence, RI: Jamestown Publishers.
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of reading. *Reading Research Quarterly, 21*, 360-406.

**Author's Note:** Correspondence may be addressed to Christopher H. Skinner at The University of Tennessee, Educational Psychology Department, Claxton Complex A-518, Knoxville, TN 37996-3400, phone (865) 974-8403, FAX (865) 974-0135, email: [cskinne1@utk.edu](mailto:cskinne1@utk.edu)

## COUNTERING TEACHER RESISTANCE IN BEHAVIORAL CONSULTATION: RECOMMENDATIONS FOR THE SCHOOL-BASED CONSULTANT

Tonya S. Butler, Adam D. Weaver, R. Anthony Doggett, & T. Steuart Watson  
Mississippi State University

Psychologists working in schools have demonstrated the effectiveness of behavioral interventions for improving academic performance and classroom behaviors. Furthermore, the amendments to IDEA have given school-based professionals greater responsibilities for conducting functional behavioral assessments and implementing positive behavioral interventions in the classroom. Typically, psychologists working in the schools use behavioral consultation as a framework for delivering behavioral services. Despite the existence of a number of studies supporting the effectiveness of delivering intervention services through a consultation framework, many applied personnel are often met with resistance from consultees that weakens the effectiveness of the intervention. As such, the present article reviews potential reasons for resistance and provides the consultant with verbal tools for effectively responding to the resistance from consultees who are asked to implement behavioral strategies. Specific exemplars of verbal resistance statements are examined along with potential responses that can be utilized by the consultant for countering the consultee's statements. The article concludes with suggestions for future research.

---

Psychologists working in the schools have demonstrated the effectiveness of behavioral interventions for improving academic performance and classroom behaviors. In addition, amendments to IDEA have given school psychologists greater responsibilities for conducting functional behavioral assessments and implementing positive behavioral interventions in the classroom (Drasgow & Yell, 2001; Gresham, Watson, & Skinner, 2001). In many instances, psychologists working in the schools use consultation as a framework for delivering behavioral services. Although there are different models of consultation from which psychologists may choose (e.g., mental health, behavioral, ecological, problem-solving, organizational), surveys of practicing school psychologists consistently indicate that the behavioral consultation model is the most consistently preferred (Costenbader, Schwartz, & Petrix, 1992; Gutkin & Curtis, 1990; Reschly & Wilson, 1995). In addition, because functional assessments, which are derived from applied behavior analysis, are mandated by IDEA for use in the schools in certain circumstances, behavioral consultation is the logical choice for delivering interventions that are behavior analytic in nature. As a result of the success of using behavioral interventions and functional methodology in school settings, consultation has become a major role of school psychologists (Fagan & Wise, 2000).

Briefly, behavioral consultation (BC) is a service delivery model that is based on a triadic relationship between a psychologist, teacher, and student. It differs from other service delivery models in that the psychologist (consultant) works with the teacher (consultee) to effect change in a student (client). BC was originally designed to enable psychologists to reach a greater number of children, to replace the traditional refer-assess-place model, and to offer more effective psychological services in the schools (Bergan, 1977; 1995). Perhaps one of the most important goals of behavioral consultation is to provide the consultee with skills that can be used efficiently with future situations such that assistance from the consultant is either unnecessary or only minimal.

Despite the existence of a number of studies supporting the effectiveness of delivering intervention services through a behavioral consultation framework (e.g., Gresham & Kendell, 1987; Gutkin & Conoley, 1990; Noell & Witt, 1998), psychologists working in the schools are often met with resistance from consultees (Hyman, Winchell, & Tillman, 2001; Tingstrom & Edwards, 1989; Witt, 1986). According to Piersel and Gutkin (1983), resistance is the result of the consultants' inability to convince the consultee to participate in the problem solving process. Dougherty (2000) referred to resistance as the failure of the consultee to participate constructively in the consultation process. Many times, this



resistance is covert and results in poor treatment integrity (Gresham, MacMillan, Beebe-Frankenburger, & Bocian, 2000; Sterling-Turner, Watson, Wildmon, Watkins, & Little, 2001). At other times, this resistance may be overt, that is verbal in nature. Teachers and other consultees are sometimes often quick to point out why an intervention will not work or is not fair to the rest of the students. Regardless of its form, consultee resistance is a topic of great concern to consultants because it may lead to the teachers' refusal to implement the procedures required for an effective behavioral intervention. As Tingstrom and Edwards pointed out:

“The fact that behavioral interventions have been proven to work in the classroom on the basis of various empirical investigations in no way ensures their implementation by teachers. Teachers must view behavioral interventions as sufficiently acceptable to be willing to implement them. The most useful and potentially most effective interventions are rendered useless if not implemented” (p. 195).

Gresham (1989) noted that one of the most critical variables in consultation is the degree to which consultees implement treatments as they are intended. It is impossible to evaluate the effectiveness, or ineffectiveness, of the intervention if it is not implemented correctly (i.e., accurately and consistently) (Watson & Robinson, 1996).

#### BASIS FOR CONSULTEE RESISTANCE

Perhaps one reason that teachers often resist consultation and behavioral interventions is their training programs often neglect to expose them to in-depth treatment of classroom management and to the consultation process that is so familiar to psychologists working in the schools (Watson, 1994). Because of their unfamiliarity with behaviorally oriented interventions and with the process of working collaboratively with a consultant, teachers may respond to these novel situations by being resistant (Bonardi, Hall, & Guthrie, 1991; Frank & Raudenbush, 1998; Harmon-Jones & Allen, 2001; Uehara, 2000).

Another possible reason for resistance during consultation is that many teachers turn to consultants for their expertise in managing child

behavior problems and knowledge of academic remediation techniques. Problems may be encountered when the consultant is unable to supply a quick remedy to the problem (Thomas & Grimes, 1997). It may be, for example, that more assessment is needed to determine if the problem is the result of a skill deficit or a performance deficit that results in an increase in the amount of time the teacher must deal with the problem before solutions are identified.

A third possible basis for consultee resistance is that consultants are often under the assumption that the consultee is entering the relationship voluntarily and is willing to make appropriate changes. Many times, however, this is not the case as the teacher has been strongly persuaded by an administrator to seek assistance with a particular student or students. On the other hand, although a consultee may voluntarily seek consultation, their expectation that they will not have to substantially change their own behavior to effect change in the student is not fulfilled and is, in most cases, the exact opposite. That is, consultants working from a behavioral perspective almost always change some aspect of the teacher's behavior in order to change student behavior. Having to change their own behavior is sufficiently aversive for some teachers that they do not actively participate in consultation although they are physically present.

Regardless of the reason for their resistance, it has been our experience that consultees are most likely to exhibit initial resistance in the form of verbal behavior. The successful consultant will be able to counter this resistance and move forward through the consultation process. Although it is important for successful behavioral consultants to correctly identify the problem, conduct an analysis, and provide appropriate plan implementation, these skills are not sufficient for effective consultant-consultee interactions. Gutkin and Conoley (1990) stressed the importance of using their interpersonal influence with the consultee before trying to address the problems presented by the client. They termed this the “Paradox of School Psychology,” because of the importance of interpersonal influence on *adults* as the key to successful services for the child. The authors further believe that it may be helpful to arm

consultants with a set of responses with which to counter common objections raised by teachers when asked to implement behavioral interventions. These responses are not intended as academic or philosophical justifications for the use of behavioral strategies but instead are to serve as exemplars of ways to respond to typical consultee objections.

### RESPONDING TO RESISTANCE

A number of related skills are important to extract the most benefit from the consultation interaction. First and foremost, it is important to remember that there is no substitute for effectiveness. Effectiveness in this context refers to expertise in utilizing the behavioral consultation model, interacting with the consultee, and in applying behavioral interventions in school-based settings. Although there are other variables that are integral to effective consultation (Dougherty, 2000), we are going to focus on one: addressing consultee verbal resistance. It is important for consultants to have skills to address verbal resistance because being unprepared to address a negative statement made by the consultee can jeopardize the perceived professional competency of the consultant (Gutkin & Curtis, 1990; Thomas & Grimes, 1997).

The purpose of this article is to provide behavioral consultants with verbal tools to effectively respond to verbal resistance from teachers and other consultees who are asked to carry out behavioral strategies. Previous articles have provided ample justification for behavioral techniques in the face of common teacher misconceptions (Tingstrom & Edwards, 1989), and strategies to effectively identify the types of resistance a consultant may encounter (Margolis, Fish & Wepner, 1990; Randolph & Graun, 1988; Witt, 1986). Although the astute consultant can easily acknowledge verbal resistance offered by the consultee, they may not have the answers in their own verbal repertoire to effectively respond to that resistance. For this reason, it is important to have readily appropriate verbal responses that will facilitate the consultee's cooperation and that will not contain language and terminology that is confusing, obscure or aversive to the consultee. Although a behavioral intervention may be

“common sense” to a consultant who has been trained in such techniques and has witnessed its effectiveness many times, many teachers may be resistant to new ideas that run counter to their training and experience.

The responses provided in this article are not intended to replace competence in problem identification, applied behavior analysis, and effective plan implementation. Instead, these responses are designed to provide exemplars for the consultant to use when faced with verbal resistance from the consultee. If the consultant is unable to promote a positive consultant-consultee relationship and is unable to establish the consultee's trust and willingness to cooperate and collaborate during the first stages, it is doubtful that an effective relationship will develop (Thomas & Grimes, 1997). The consultant-consultee relationship is considered to be a vitally important variable in the outcome of consultation, regardless of the specific model being used (Dougherty, 2000; Gutkin & Curtis, 1990).

### EXEMPLARS OF VERBAL COUNTERRESISTANCE TECHNIQUES

Listed below are some of the most common verbal resistance statements offered by teachers during behavioral consultation. For each teacher resistance statement, we have offered several options for the consultant's consideration when responding.

**Teacher Response:** It is not fair to the rest of the class to provide special rewards to one student.

**Option 1:** You could give the reward to the child in private. This would also provide one-on-one attention to the target child. However, if you do not feel comfortable with this then we can implement the intervention for the whole classroom.

**Option 2:** Since the child is disruptive to the class, the other students would probably appreciate an intervention to decrease the disruptive behavior. The other students are probably well aware of the fact that this child needs extra help. However, if it does get to be a problem then I would simply remind the

students that each child in this room is provided with their own individual needs.

**Option 3:** Lets think about children with academic problems. We are usually not concerned about developing a special program to assist students who have not mastered certain academic skills as well as their peers. For example, we typically only provide tutoring to the individual student that is having problems in reading as opposed to setting up a comprehensive reading intervention for the whole class.

**Teacher Response:** I shouldn't have to reinforce or reward a student for doing something that they should be doing anyway.

**Option 1:** You're probably right, but since the child is acting inappropriately we need to focus on providing him with opportunities to learn the behaviors that we desire from him/her. We have to remember that this student may not have learned the difference between inappropriate and appropriate behaviors. If you only reward the behaviors that you desire from the child, the child will soon catch onto which behaviors are providing reinforcement.

**Option 2:** We discussed earlier that you thought that this student enjoyed it when you were correcting him. Now, rather than reinforcing his inappropriate behaviors you will be providing reinforcement for his/her appropriate behaviors.

**Option 3:** Everyone needs some type of reinforcement for his/her behavior. We aren't here today simply because we enjoy our jobs. While many of us obtain great personal satisfaction from our jobs, we also expect to receive a paycheck for our efforts. Children are not any different from most adults in that respect when you think about it.

**Option 4:** I agree with you in principle. You shouldn't have to, and actually, you don't have to. But if this child's behavior is as much of a problem as you say it is, then it will take a little bit of extra effort until it is under control.

**Teacher Response:** I don't have the time to deliver reinforcers or pay attention when the target child is behaving appropriately.

**Option 1:** You mentioned earlier that you are tired of spending so much of your time on trying to resolve this child's behavior problems. Now, rather than spending your time on correcting, redirecting, punishing, and explaining how much the child disrupts your class during parent-teacher conferences you can use it for reinforcing the child during critical learning opportunities.

**Option 2:** The target child cannot expect to have all of your time. But we can come up with some ways of giving him/her brief frequent bursts of attention that only uses a small portion of your time (e.g. pat on head, allow him/her to pass out papers, allow him/her to be the line leader).

**Option 3:** Provide the teacher with a conditional probability chart demonstrating the amount of time that the disruptive behavior is followed by social attention versus the amount of time that appropriate behavior is followed by the same consequence.

**Teacher Response:** There is nothing we can do at school to change this child's behavior. The problem stems from his home life.

**Option 1:** The child's behavior may stem from his home life, however our goal is to control the child's behavior during school hours. Since we know we cannot control the child's home life, we need to focus on how to structure a safe and reinforcing environment for the child here at school.

**Option 2:** The home situation is terrible. But all we can control are these seven hours, probably the best and most consistent seven hours of this child's day. Children quickly learn the settings in which they can perform certain behaviors. (Children know that it okay to run and yell at recess, but they also know that those behaviors are not appropriate in a church.)

**Option 3:** If the home life is terrible, then the job of the teacher is even more important because this is probably the only

chance the child will have to learn appropriate skills.

**Teacher Response:** Why should the child be reinforced for one good behavior when they have been misbehaving all day?

**Option 1:** You are correct; you do not want the child to think he/she is being rewarded for acting inappropriately. Therefore, we need to ensure that the child is reinforced for displaying appropriate behavior more often than he or she is reinforced for displaying inappropriate behavior.

**Option 2:** If you reinforce the appropriate behavior consistently, the child will learn that he/she is only getting rewarded when he/she acts appropriately.

**Teacher Response:** Reinforcement or punishment just does not work with this child; I have tried everything. There must be something else happening.

**Option 1:** It is important to remember that what works for most children may not always work for the child with special needs. We need to focus on identifying the activities that are reinforcing for this individual child.

**Option 2:** By definition, reinforcement increases the probability that a behavior will occur, therefore, if he/she is continuing to perform the behavior, then it is being reinforced.

**Teacher Response:** The student is supposed to get on medication in the next week or so. I am not going to do anything yet. I am going to see what effects the medication is going to have.

**Option 1:** Medication may be effective, but what are you going to do when the child forgets his/her medication? If we teach the child appropriate behavior for the classroom then you will not have to worry about whether the child took his/her medication this morning.

**Option 2:** Children cannot be on medication for long periods of time. Additionally, medication does not teach children the skills that they need to know. So when the

child is taken off his/her medication you will not have to come up with an intervention to decrease the inappropriate behaviors that will emerge once medication is ceased.

**Option 3:** Medication is not always effective, so you need to develop a plan to manage the occurrence of problem behavior that is not based solely on the use of medication alone.

**Teacher Response:** Shouldn't we focus more on punishing the bad behavior than trying to focus on good behavior? "Catching them being good" never works.

**Option 1:** If we only punish inappropriate behavior then we are not providing the child with an opportunity to learn replacement behaviors. Not only do we want to take this chance to decrease the child's inappropriate behaviors, but we also want to replace those behaviors with appropriate ones.

**Option 2:** If we only focus on the child when he/she is acting inappropriately this tells the child that he/she will only receive attention when they are engaging in problem behavior. Children would rather have negative attention than no attention at all.

**Option 3:** We don't want the child to associate you with punishment only. If we focus on using reinforcement to increase the occurrence of appropriate behavior, then the child will soon pair you with the use of reinforcement.

**Teacher Response:** This may work for younger children but it is not going to work for older children.

**Option 1:** Once you identify the environmental events that are causing the behavior to occur, you can change the events in the environment in order to change anyone's behavior. Now our job is to identify appropriate ways to alter environmental events so as to influence the behavior.

**Option 2:** We can provide environmental contrast for a child of any age. For example, when I was observing I noticed

that each time the target child disrupted the class you would verbally reprimand him. We can use this as a chance to withdraw your attention from the child when he/she is being disruptive. However, when he/she is acting appropriately you can verbally praise him/her. This provides the child with the high contrast needed to help him/her learn which behaviors earn reinforcement and which behaviors do not. Even adults quickly pick up on environmental contrast. We know if we don't come to work, we don't get paid. However, if we do come to work we do get paid.

**Teacher Response:** This behavior stuff is just common sense.

**Option 1:** In a lot of ways, these principles are very straightforward. However, if not implemented in a very systematic manner, then the child often still does not learn how to follow instructions or classroom rules. Understanding the principles should be easy, carrying them out in a consistent manner may be more difficult, and require more work.

**Teacher Response:** I have been teaching for 20 years and I can tell you right now that this is not going to work.

**Option 1:** A lot has changed in 20 years. I would like me to show you some procedures that may make your life easier and improve the child's behavior.

**Option 2:** Maybe not, however, I have seen these procedures work many times. You mentioned that nothing you have tried has worked so maybe we should give this intervention a chance.

**Teacher Response:** Obviously, you have never been a teacher.

**Option 1:** You're correct. I never have. That's why I'm relying on your expertise in this area. If we utilize both your expertise in this area and my expertise in behavioral management, I'm sure we can be successful in managing the child's behavior.

**Option 2:** Actually, I do have experience as a teacher, and I can completely

understand your feelings toward this child and the difficulties involved. But I know that you didn't become a teacher because it was an easy job. It's times like this that allow you to really make a difference in a child's life, and I know that's probably one of the primary reasons you became a teacher.

## DISCUSSION

There have been many attempts to provide data for the reasons that teacher resistance occurs and to identify typical types of resistance (Abidin, 1975; Piersel & Gutkin, 1983; Witt, 1986). A limitation to this research was that the consultant was not provided with tools to overcome the resistance. Our goal is to provide the consultant with responses to common types of challenging statements made by teachers during the planning stage of behavior consultation. After interviewing consultants who work in the schools, the authors took the statements that the consultants were typically faced with when an intervention was proposed and then paired them with responses that have been effective in promoting a successful intervention based on prior clinical experience.

Future research needs to systematically address how teachers typically react to the aforementioned responses. During the early stages, this type of research will be conducted in an analogue situations allowing researchers to identify the verbal responses that are most likely to counter teacher resistance often experienced during the initial stages of the consultation relationship. Once the more favored responses were identified, more applied research would ensue with actual consultants using the responses during naturalistic consultation episodes. While not exhaustive as a list of possibilities for measuring potential outcomes, data on initial acceptance of the responses, continued active resistance, and/or passive resistance to implementing the techniques with integrity would be evaluated at a minimum. Without furthering understanding of interpersonal influences on consultees, the effort toward creating effective interventions that are implemented with high integrity will go unfulfilled. However, continued research on consultee resistance and the consultant behaviors

utilized to eliminate or reduce that resistance will allow the field to further evaluate the stimuli that promote positive consultation relationships which ultimately lead to effective interventions for children and teachers.

## REFERENCES

- Abidin, R. A. (1975). Negative effects of behavioral consultation: "I know I ought to, but it hurts too much." *Journal of School Psychology, 13*, 51-57.
- Bergan, J. R. (1977). *Behavioral consultation*. Columbus, OH: Merrill.
- Bergan, J. R. (1995). Evolution of a problem-solving model of consultation. *Journal of Educational and Psychological Consultation, 6*, 111-123.
- Bonardi, C., Hall, G., & Guthrie, D. (1991). The effect of a retention interval on habituation of the neophobic response. *Animal Learning & Behavior, 19*(1), 11-17.
- Costenbader, V., Swartz, J., & Petrix, L. (1992). Consultation in the schools: The relationship between preservice training, perception of consultative skills, and actual time spent in consultation. In T. S. Watson, K. P. Ray, H. Sterling Turner and Patricia Logan (1999). Teacher-implemented functional analysis and treatment: A method for linking assessment to intervention. *School Psychology Review, 28*, 292-303.
- Dougherty, A.M. (2000). *Psychological consultation and collaboration in school and community settings* (3<sup>rd</sup> ed.). Belmont, CA: Wasworth/Thomson Learning.
- Drasgow, E. & Yell, M. L. (2001). Functional behavioral assessments: Legal requirements and challenges. *School Psychology Review, 30*, 239-251.
- Fagan, T. K. & Wise, P. S. (2000). *School psychology: Past, present, and future* (2<sup>nd</sup> ed.). Bethesda, MD: NASP.
- Frank, R.A., & Raudenbush, B. (1998). Individual differences in approach to novelty: The case of human food neophobia. In R.R. Hoffman, M.F. Sherrick, et al. (Eds.), *Viewing psychology as a whole: The integrative science of William N. Dember*. (pp. 227-245). Washington, DC: American Psychological Association.
- Gresham, F. M. (1989). Assessment of treatment integrity in school consultation and prereferral intervention. *School Psychology Review, 18*, 37-50.
- Gresham, F. M. & Kendall, G. K. (1987). School consultation research: Methodological critique and future research directions. *School Psychology Review, 16*, 306-316.
- Gresham, F. M., MacMillan, D. L., Beebe-Frankenberger, M. E., & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatment is implemented? *Learning Disabilities Research & Practice, 15*, 198-205.
- Gresham, F., Watson, T. S., & Skinner, C. H. (2001). Functional behavioral assessment: Principles, procedures, and future directions. *School Psychology Review, 30*, 156-172.
- Gutkin, T. B., & Conoley, J. C. (1990). Reconceptualizing school psychology from a service delivery perspective: Implications for practice, training, and research. *Journal of School Psychology, 28*, 203-223.
- Gutkin, T. B., & Curtis, M. J. (1990). School-based consultation: Theory, techniques, and research. In T. B. Gutkin & C. R. Reynolds (Eds.), *The handbook of school psychology* (2nd ed.), (pp. 577-611). New York: John Wiley & Sons.
- Harmon-Jones, E., & Allen, J.B. (2001). The role of affect in the mere exposure effect: Evidence from psychophysiological and individual differences approaches. *Personality and Social Psychology Bulletin, 27*, 889-898.
- Hyman, I. A., Winchell, K., & Tillman, T. (2001). Treatment integrity and satisfaction using the teacher variance approach: A multidimensional method for dealing with teacher resistance. Presented at the Annual Convention of the National Association of School Psychologists, Washington, D.C.
- Margolis, H., Fish, M., & Wepner, S. (1990). Overcoming resistance to prereferral classroom interventions. *Special Services in the School, 6*, 167-185.
- Noell, G. H., & Witt, J. C. (1998). Toward a behavior analytic approach to consultation. In T. S. Watson & F. M. Gresham (Eds.), *Handbook of child behavior therapy: Issues in clinical child psychology* (pp. 41-57). New York: Plenum Press.
- Piersel, W. C. & Gutkin, F. B. (1983). Resistance to school-based consultation: A behavior analysis of the problem. *Psychology in the Schools, 20*, 311-320.
- Randolph, D. L., & Graun, K. (1988). Resistance for consultation: A synthesis for counselor consultants. *Journal of Counseling and Development, 67*, 182-184.
- Reschly, D. J., Wilson, M. (1995). School psychology practitioners and faculty: 1986 to 1991-92 trends in demographics, roles, satisfaction, and system reform. In T. S. Watson, K. P. Ray, H. Sterling Turner and Patricia Logan (1999). Teacher-implemented functional analysis and treatment: A method for linking assessment to intervention. *School Psychology Review, 28*, 292-303.
- Sterling-Turner, H. E., Watson, T. S., Wildmon, M., Watkins, C., & Little, E. (2001). Investigating the relationship between training type and treatment integrity. *School Psychology Quarterly, 16*, 56-67.
- Thomas, A., & Grimes, J. (Eds.) (1997). *Best practices in school psychology-III*. Bethesda, MD: National Association of School Psychologists.
- Tingstrom, D. H., & Edwards, R. (1989). Eliminating common misconceptions about behavioral psychology: One step toward increased academic productivity. *Psychology in the Schools, 26*, 194-202.
- Uehara, I. (2000). Transition from novelty to familiarity preference depending on recognition performance by 4-yr-olds. *Psychological Reports, 87*(3), 837-848.
- Watson, T. S. (1994). The role of preservice education in training teachers to serve behaviorally disordered children. *Contemporary Education, 65*, 128-131.
- Watson, T. S., & Robinson, S. L. (1996). Direct behavioral consultation: An alternative to traditional behavioral consultation. *School Psychology Quarterly, 11*, 267-278.
- Witt, J. C. (1986). Teachers' resistance to the use of school-based interventions. *Journal of School Psychology, 24*, 37-44.

**DECREASING THE INTENSITY OF REINFORCEMENT-BASED  
INTERVENTIONS FOR REDUCING BEHAVIOR: CONCEPTUAL ISSUES  
AND A PROPOSED MODEL FOR CLINICAL PRACTICE**

Linda A. LeBlanc  
Western Michigan University  
Louis P. Hagopian  
Johns Hopkins University School of Medicine  
Kristen A. Maglieri  
Kennedy Krieger Institute  
Alan Poling  
Western Michigan University

Behavioral interventions that include reinforcement as a treatment component have proven quite effective in decreasing problem behavior in children and individuals with developmental disabilities. These interventions are typically initiated with frequent, immediate reinforcement to increase the likelihood of success and schedules may then be thinned to more clinically manageable schedules to promote generalization and maintenance of treatment effects. Immediate reinforcement can also be delayed to the same effect. However, there are currently no specific procedural guidelines for decreasing the intensity of effective behavioral interventions. The current paper examines several conceptual issues regarding procedures for decreasing the intensity of behavioral interventions and presents clinical and research suggestions.

---

Behavioral interventions have proven quite effective in decreasing problem behaviors of children, especially those with developmental disabilities (Carr, Coriaty, & Dozier, 2000; Carr, Yarbrough, & Langdon 1997; Watson & Gresham, 1998). Many behavioral interventions used to reduce behavior incorporate reinforcement as a treatment component in the form of differential reinforcement of appropriate behavior (DRA), differential reinforcement of other behavior (DRO) or noncontingent reinforcement (NCR) (Carr, Coriaty, Wilder et al., 2000; LeBlanc, Le, & Carpenter, 2000). In an attempt to increase the likelihood of success, such interventions typically involve frequent and immediate delivery of high-quality reinforcers when treatment is initiated (Hagopian, Fisher, & Legacy, 1994; Miltenberger, 1997). Once the intervention has proven effective, the intensity of the intervention is gradually decreased over time by either reducing the frequency of reinforcement or implementing a delay to reinforcement.

Interventions of decreased intensity are beneficial for several reasons. First, frequent, and immediate delivery of reinforcers often is unmanageable in natural settings (Fisher et al, 2000). For example, a parent may not be able to provide continuous attention for their child during each visit to the doctor or during an important telephone conversation. Second, less

labor intensive interventions may result in greater overall treatment integrity. Interventions that are labor intensive or attract undue attention in natural settings are less likely to be implemented consistently than interventions that appear more natural and are easier to implement. Third, less intensive interventions characteristically are more similar to naturally occurring contingencies than highly intensive interventions, promoting generalization of treatment effects to everyday settings (Stokes & Baer, 1977).

Unfortunately, several problems may arise when attempting to decrease the intensity of behavioral interventions. First, treatment gains may be lost as interventions become less intensive. For example, Hagopian, Fisher, Sullivan, Acquisto, and LeBlanc (1998) found that effective functional communication training with extinction interventions failed to remain effective in 60% of cases when schedule thinning or delay to reinforcement was initiated. Second, there are no accepted algorithms or even general strategies that identify specifically how the intensity of interventions should be reduced. The intensity of interventions can be manipulated along several dimensions, including the frequency of reinforcement, the delay of reinforcement, and the characteristics (e.g., magnitude, quality) of the reinforcing events. Researchers have failed to compare the effects

of altering these dimensions. Most studies that have systematically examined strategies for reducing the intensity of reinforcement-based interventions for reducing behavior have manipulated the frequency of reinforcement (Lalli, Casey & Kates, 1997; Piazza, Moes, & Fisher, 1996). Even in this area, little evidence exists to guide how quickly we should decrease intensity, or even what the initial intensity of the intervention (i.e., rate of reinforcement) should be. The research that exists has primarily been conducted under highly controlled analogue conditions, making it somewhat difficult to translate findings into useful clinical guidelines for practitioners. Rather than clear guidelines that promote systematic, effective, and efficient methods for decreasing intervention intensity, common lore and best guesses are frequently used in clinical settings.

The purpose of the present manuscript is to consider some issues relevant to reducing the intensity of reinforcement-based interventions and to offer some suggestions for doing so in applied settings. First, we describe clinical variations of schedule thinning and delaying reinforcement and their corresponding behavioral processes. Second, we briefly review the existing literature on efficacy of methods for decreasing the intensity of behavioral interventions. Third, we provide a proposed model for determining how quickly progress can be achieved in individual cases. Finally, we offer suggestions for further research. Because several studies have examined procedures used to reduce problem behaviors in children and individuals with disabilities, attention will be focused on this area. Many of the general issues discussed in this context also are relevant to reducing the intensity of reinforcement-based procedures intended solely to increase appropriate responding.

#### PROCEDURES FOR DECREASING INTERVENTION INTENSITY

To the best of our knowledge, there are no published formal decision rules for decreasing intensity. The result is that many well-intentioned behavior analysts are left without clear plans for quickly reducing the intensity of their interventions to a level that is appropriate for clients' everyday environments.

Thus, many clinical behavior analysts must rely solely on clinical lore and intuition or do not decrease intensity at all. When intensity is decreased, the progression is usually slow and deals solely with frequency of reinforcement. In actuality, as discussed in the next section, interventions may be decreased in intensity along several dimensions. Distinguishing among these dimensions is of practical as well as theoretical significance, and is made easier by use of accurate and consistent terminology.

The term "fading" is frequently used to refer to all procedures for making interventions less intensive and more manageable (e.g. schedule thinning, delaying reinforcement, prompt reduction). This term is appropriate when used to refer to gradual reductions in the frequency or intensity of prompts or other antecedent stimuli, but it is inappropriate when used to describe decreases in the frequency of reinforcement (or punishment) alone (Martin & Pear, 1999; Miltenberger, 1997). "Fading" also is inappropriate when used to refer to procedures that involve increasing delay to reinforcement (or punishment) in the absence of antecedent stimulus manipulations.

"Schedule thinning," or "thinning," is an appropriate term for referring to decreases in the frequency of reinforcement (or punishment) (Cooper, Heron & Heward, 1987; Martin & Pear, 1999). Schedule thinning involves either an increase in the response requirements before delivery of a reinforcer or an increase in the time interval preceding delivery of a stimulus. Common practice involves thinning the schedule by a small increment, then continuing to thin the schedule as long as the intervention remains effective. For instance, a fixed-ratio 1 (FR 1) schedule of reinforcement might be thinned to FR 2, 3, 4, and so on until a terminal value, perhaps FR 25, is achieved. If problems arise, the schedule returns to a previously successful level, a manipulation frequently called "backing up," and is retained at that level until positive effects are again achieved. At that point, thinning begins again. While such practices promote slow steady progress, they do not ensure that the intervention will reach a manageable intensity as rapidly or as efficiently as possible.



A third set of procedures involves systematically increasing the delay to reinforcement. In practice, this often involves gradually introducing a delay or "wait" period between the occurrence of the target response and the delivery of the reinforcer for the response, and then increasing the delay until the procedure can be sustained in the participant's everyday environment.

Attempts to deal with problem behaviors in research settings and in naturalistic environments often combine elements of fading and schedule thinning, and some of them include increasing delay to reinforcement. Nonetheless, it is important to distinguish among the three general techniques for reducing the intensity of interventions. These techniques involve different kinds of environmental manipulations and may differ in their effects. Moreover, they may not be equally appropriate in a given situation and should not be viewed as interchangeable. The following sections deal specifically with schedule thinning manipulations, which are widely used for reducing the intensity of reinforcement-based response-reduction procedures.

### **Schedule Thinning: Differential Reinforcement of Alternative Behavior**

Under a differential-reinforcement-of-alternative-behavior schedule (DRA), reinforcers are delivered dependent on the occurrence of a response that is incompatible with the behavior targeted for reduction. When schedule thinning is incorporated into DRA interventions, an increasing number of occurrences of the targeted appropriate behavior typically must occur before the reinforcer is delivered. The schedule is typically increased from a continuous (FR 1) reinforcement schedule to an intermittent FR or variable-ratio (VR) schedule. For example, if a child exhibited problem behavior in academic work settings, the function-based intervention might involve differential reinforcement of completion of tasks or communication. The intervention might begin with every instance of task completion resulting in access to reinforcers. Gradually the child would have to comply with multiple commands or complete multiple tasks to gain access to reinforcers. Piazza et al. (1996)

demonstrate schedule thinning with a DRA procedure to treat the escape-maintained aggressive behavior of a child with autism. Schedule thinning progressed in increments of one until they reached a terminal value of FR 28. The reinforcer in this case was termination of the demand context (a negative reinforcer) and the thinning procedure was quite lengthy but effective.

Hagopian et al. (1998) illustrate another example of this type of schedule thinning with their treatment of escape-maintained problem behavior by establishing a communication response as an alternative behavior. They progressed from an FR 1 to an FR 5 schedule before switching to a VR schedule that more closely approximated contingencies in the natural environment. Under a VR schedule, on average every  $n$ th response produces the reinforcer, but the number of responses required for reinforcement varies across a prearranged series of values. For example, on average every 10<sup>th</sup> response is reinforced under a VR 10 schedule, but specific reinforcers might be delivered following as few as 1 or as many as 25 responses. The reader should note that when schedules are thinned by increasing ratio requirements, the time between reinforcers usually increases also, because it takes more time to meet the new response requirements.

Procedures that involve thinning schedules by increasing ratio requirements are best used with appropriate behaviors that you want to see occur at moderate to high and steady or increasing rates. A secondary effect might be the decrease or elimination of problematic behavior. Common targets might include completion of chores, compliance with simple directives, exercise for children with weight problems, medication compliance, and completing academic assignments.

Ratio-based thinning procedures should not be used if high rates of the reinforced behavior could cause problems, as is the case with many communication-based interventions. For example, consider a school student who exhibits escape-maintained disruptive behavior during independent seatwork. The student might be taught to engage in an alternative appropriate behavior (e.g., raising a hand to ask for help)

when asked to complete work independently. Problem behavior may decrease because hand raising now produces escape and task-related social interactions. While raising a hand to request help is a preferred alternative behavior, it is not desirable to have a child raise her or his hand at a high rate throughout independent seatwork. Here, an increasing ratio schedule in all likelihood would engender a topographically and functionally appropriate behavior that constitutes a problem because of its excessive rate. Interventions meet their doom when children ask repeatedly for things or incessantly perform any of a wide range of behaviors that are highly desirable when less frequent and are annoying when more frequent. It is indeed possible to have too much of a good thing. Additionally, children should not need to ask for help repeatedly before receiving it; however, it is reasonable to expect a child to ask once and then wait (delay to reinforcement) for that assistance while a teacher works with other children.

#### **Schedule Thinning: Time-Based Procedures**

Two common interventions that involve increasing time-based schedules are differential-reinforcement-of-other-behavior (DRO) and fixed-time (FT) schedules. In the applied literature, FT schedules are commonly referred to as noncontingent reinforcement (NCR) schedules, or enriched environments. Schedule thinning with DRO procedures involves an increasing interval of time without the occurrence of the target behavior before the reinforcer is delivered. For example, Vollmer, Iwata, Zarcone, Smith, and Mazaleski (1993) initially implemented a DRO 10-s for self-injurious behavior. That is, if 10 seconds passed without the occurrence of self-injury, a reinforcer was delivered. Upon the occurrence of self-injury, a new 10-s interval was initiated. The interval was reset each time self-injury occurred. The length of the DRO interval was gradually increased until 5 min elapsed without self-injury before the reinforcer was delivered.

In general, thinning DRO schedules is possible only in situations where an adult can provide high levels of monitoring for the behavior that is to be decreased. For example, it might be possible to reduce how often an older

sibling inappropriately took toys from a younger sibling by reinforcing gradually lengthening intervals during which toys were not stolen. This procedure would be effective only if theft could be detected accurately and DRO intervals could be accurately monitored and reset if needed.

DRO arrangements do not necessarily foster appropriate behavior and are schedules of reinforcement only in the sense of increasing intervals of specified length during which a target behavior fails to occur (Poling & Ryan, 1982). Put differently, they reinforce the *omission*, not *emission*, of behavior. Therefore, although DRO arrangements may be effective in reducing undesired behavior, they must be augmented to produce desired responding. It is noteworthy that DRO arrangements may be used effectively when a strong repertoire of appropriate behavior exists but is not often used in favor of inappropriate behaviors. For example, LeBlanc, Hagopian, and Maglieri (2000) implemented DRO for a man with inappropriate social behaviors who already had an extensive repertoire of appropriate social behaviors.

Under FT schedules (i.e., NCR procedures), a stimulus demonstrated to be a positive reinforcer in another setting is delivered at preset intervals regardless of the occurrence or non-occurrence of behavior. Thinning of FT schedules involves a gradual increase in the interval before presentation of the stimulus. There is no dependency between a target behavior and the presentation of the stimulus. A common initial schedule for NCR is continuous access or near continuous access to the reinforcer. Gradually, the schedule is thinned until stimuli are available only periodically, perhaps every 1-2 minutes or eventually every 10-20 minutes. Hagopian et al. (1994) illustrated the importance of schedule thinning in NCR with attention-maintained problem behavior of quadruplets. NCR was unsuccessful when started at a FT 5-min schedule but was effective when begun with near continuous attention (FT 10-s) gradually thinned until the schedule FT 5-min.

A sizeable number of studies have demonstrated the value of NCR in reducing

troublesome behaviors (Carr, Coriarty et al., 2000). Like DRO schedules, FT schedules do not necessarily generate appropriate responding. They appear to reduce undesired behavior through extinction and satiation.

### **Delay to Reinforcement: The "Waiting" Game**

Delay to reinforcement has most often been arranged in the context of DRA schedules. Here, the procedure involves maintaining the original FR 1 schedule of reinforcement and gradually introducing a waiting period between the time of the necessary response and the delivery of the reinforcer. Perhaps the best example is a child who makes a reasonable request that cannot be immediately met. When a child asks at 4:15 to watch an appropriate TV program scheduled for 4:30, the child will preferably ask only once before the reinforcer is delivered and will wait without problem behavior until the reinforcer can be provided. This intervention is desirable to use when you want to see relatively low but steady rates of responding and minimal emotional behavior.

Waiting is very difficult for many people, and thinning schedules by increasing delay to reinforcement is invaluable for teaching this skill. Research suggests that control of behavior with delayed reinforcement is easiest to obtain when the delay is increased very gradually, the magnitude of reinforcement is increased as a function of delay, and care is taken (where possible) to foster rule-governed behavior that is consistent with desired outcomes (e.g., Logue, 1988). Additionally, Fisher et al. (2000) suggest that inserting alternative activities into the "wait" interval facilitates tolerance of delay to reinforcement. Hanley, Iwata, and Thompson (2001) compared several procedures for increasing delays following communication-based interventions. They determined that increasing delays resulted in weakening of the new communication response while adding additional stimulus cues to signal the wait interval resulted in more rapid schedule thinning, maintenance of the new communication response, and low rates of problem behavior.

### **Procedures for Determining Progression of Schedule Thinning**

Knowing which procedures to use to decrease intervention intensity is a first step, but you must also know how much the intensity can be decreased and how quickly this can occur. An intervention that begins at a low intensity may never prove effective and an effective intervention that changes too rapidly can fall apart. There are two typical methods for setting the initial level of intensity and the speed of progression described in intervention research. The first method uses pre-set values and the second method involves mathematical calculations based on ongoing child behavior. Perhaps the most common method is to use the most intense intervention possible (e.g., FR 1, DRO 1-s) and to use small preset increments of increase. This method provides the greatest assurance that the intervention will prove effective, however, it typically presents a longer road to the terminal schedule and may be quite time consuming. As discussed previously, Piazza et al. (1996) illustrated this method by proceeding from an FR 1 to an FR 28 schedule in increments of 1.

Hagopian et al. (1998) illustrate the use of steady slow progression with delay to reinforcement procedures with an individual with problem behavior maintained by access to tangible reinforcers. The initial intervention involved reinforcement for a communication response on an FR 1 schedule with no delay between the communication response and the delivery of the reinforcer. As schedule thinning was implemented, a 5-s delay was inserted between the communication response and the delivery of the reinforcer. This delay was increased by 5-s intervals fourteen times (i.e., to 95 s) and then by 10 s, 15 s, and 30 s until the terminal delay was reached (300 s). With particularly severe and difficult cases, such a slow progression may be necessary to ensure continued success, however, for most cases encountered in outpatient settings procedures that move faster can be used.

The second method involves setting the initial interval based on the average inter-response time (IRT) or rate of behavior during baseline observations and gradually adjusting

the schedule based on the average for the most recent intervention sessions. This procedure allows the intervention to begin at a much less intense level, potentially facilitating later decreases in intensity. The drawback is that this method requires constant mathematical calculations and adjustments. These calculations, while easily conducted in laboratory and analogue settings, may prove cumbersome for typical clinicians working with children and their families in outpatient settings. Also, depending on actual IRT distributions, procedures with parameters based on *average* IRTs may prove ineffective, and analyses that are more sophisticated may be required to establish effective treatment parameters. A related procedure for setting the initial would involve calculating the average latency and decreasing that schedule by 50% to create a slightly higher probability of success.

Lalli et al. (1997) illustrated a combination of these two procedures using FT schedules with three children with problem behavior maintained by access to tangible reinforcers. They set their initial level of intensity based on averages during baseline, then systematically increased the interval between reinforcer deliveries by a pre-set interval (i.e., 30 s or 120 s, depending on the child involved) until the terminal level was reached for each individual. They used a criterion of two consecutive successful sessions (e.g., no or little problem behavior) before the schedule was progressed. They also returned to the previous schedule when three consecutive sessions were unsuccessful (i.e., problem behavior equal to or exceeding baseline).

Kahng, Iwata, DeLeon, and Wallace (2000) directly compared procedures for programming FT schedules. In one condition, an arbitrarily selected dense schedule of reinforcement was gradually increased by a fixed time increment. In the second condition, the initial reinforcement schedule was determined by calculating the mean IRT for baseline sessions and gradually adjusting the schedule based on the mean IRT for the most recent intervention sessions. Both schedules were determined to be effective though the adjusting IRT method resulted in slightly

quicker progress to the terminal reinforcement schedule.

LeBlanc, Hagopian and colleagues have developed a slightly different procedure (LeBlanc et al., 2000; LeBlanc, Hagopian, Marhefka, & Wilke, 2001) that may prove easily modified for general outpatient clinical use. The procedure is based on a mathematical model with proportional increments in schedules. A series of values, referred to as steps, is established which generally represents a 33 to 50 % increase from the previous value. Early steps generally represent a 50% increase, while later steps represent a 30-40% increase and can be rounded to an easily remembered value (e.g., 115 seconds rounded to 2 minutes). See Table 1 for a sample series of steps that might be used when using an increasing ratio schedule and a sample series of steps for a DRO schedule. The proportional increments allow one to avoid the problems that can occur with a simple "doubling" procedure. Doubling a schedule value is reasonable early in schedule thinning (e.g. FR 1 to FR 2, FR 2 to FR 4), but quickly leads to unworkably large values (e.g. FR 8 to FR 16, FR 16 to FR 32). With the procedure described by LeBlanc et al. (2000; 2001), the steps proceed until the terminal schedule value is achieved. The criterion for increasing the schedule is a 90% or greater reduction from baseline for two consecutive sessions. If problem behavior does not remain at a 90% reduction for two consecutive sessions, the schedule returns to the most recent successful level of intensity.

LeBlanc et al. (2000) demonstrated use of these steps for an increasing DRO schedule in their treatment of inappropriate social behavior in an adult with mental retardation. The initial DRO schedule was set at the mean inter-response time (IRT) during the baseline phase. The initial interval (20 s) was entered as the starting point and nine subsequent steps were established. The initial thinning (step 1) represented a 33% increase in the duration of the DRO interval and the length of the session (600 s) was set as the DRO interval for step 9. All increases ranged from 33% to 100% with an average increase of 45%. At each step, two sessions were required to meet the criterion for success before the next step was implemented.

### Recommendations and a Proposed Model

The reader should note that the following recommendations are based on the relevant research and our applied experience; however, this proposed model has not been directly tested in multiple studies. We encourage readers to consider the practices delineated as recommendations based on research and clinical experience rather than documented best practice. In addition, we invite independent investigators and practitioners to evaluate the utility and efficiency of the model or to respond with alternative models.

We make the following recommendations for decreasing intervention intensity when using reinforcement-based procedures to treat problem behavior. First, begin the intervention at a relatively intensive level to promote initial success, a proposal that is supported by Hagopian et al (1994). The initial schedule can be determined by a) selecting the most intense schedule manageable (e.g., Fr-1, 1-s to 5-s delay) or b) determining the schedule by calculating an average IRT or success level from baseline observations (Kahng et al, 2000). Second, determine which method of schedule thinning is appropriate for your intervention. Increasing ratio schedules are valuable when you want to see high rates of behavior (e.g., correctly completed math problems), whereas delay to reinforcement is preferable when you want a behavior to occur once followed by a reasonable wait before the reinforcer is delivered. Third, determine a reasonable terminal goal for the intensity of your intervention based on the age and functioning level of the child. For example, a 10-20 minute wait may be reasonable for a 8-year-old child while a 5 minute wait is appropriate for a younger child or a child with developmental delays. To enhance social validity, determine the terminal schedule value in consultation with the individual who ultimately will use the intervention in a natural setting.

Fourth, establish a series of values or steps between the initial level and proposed terminal level. We recommend approximately 8-12 steps, but the number will vary according to the difference between the initial level and terminal level. Increases of approximately 33 to

50 % from the previous schedule are reasonable, with the initial increments larger than the later increments. For procedures that begin at the highest intensity, the initial steps may have to be 100% increments (e.g., FR 1 to FR 2). These increments avoid the problem of excessive increases associated with doubling procedures and allow calculation of values that are useful in natural environments. Our recommendation includes flexibility to accommodate the purposes of both researchers and clinicians. For example, a researcher in a laboratory setting might choose to set a specific increment (e.g., 50%) throughout all steps in order to specifically investigate the effects of increment magnitude even though 50% increases might create unusual schedule values (e.g., 157-s delay). Clinicians might opt to go with values such as those published here, which have reasonably varying increment values but are easy for parents or teachers to remember (e.g., 30-s, 2-min).

Fifth, allow sufficient exposure to the new level of intervention before proceeding to the next level. We have found that two consecutive exposures during which performance is at an acceptable level is a useful criterion for defining sufficient exposure (LeBlanc et al, 2000; 2001), however this recommendation has not been empirically tested. One exposure would allow quicker progression while additional exposures might be beneficial for certain interventions procedures where greater behavioral variability might be expected (e.g., extinction). Sixth, if success is not achieved at a new level after multiple exposures, briefly return to a previously successful level before trying again. In most cases, this should occur following no more than five unsuccessful exposures. Again, this recommendation has not been empirically compared to other procedures, but has proven useful in clinical practice.

Finally, progress as quickly as possible to the final intervention level by using periodic probes to determine how quickly the intensity reduction can be advanced. After three consecutive steps are successfully completed (e.g., initial level, step 1, step 2, step 3), probe three steps higher (e.g., step 6). If the probe is unsuccessful, return to the most recently successful step (e.g., step 3) and begin slow progression through the three intermediate steps

from that point before conducting the next probe (e.g., step 9). If the probe session is successful, immediately probe three additional steps higher. When the progression has reached the halfway point (e.g., 5-minute wait for a goal of 10 minutes), you may choose to probe at the terminal schedule value. We do not recommend conducting probes at the terminal-step prior to the halfway mark because the differences between current schedule and terminal schedule are typically extremely salient prior to that point. See Table 2 for a sample probe procedure. Although the study described below successfully used this model, replications of the model have not yet been published and we encourage readers to evaluate this procedure empirically, as we are currently doing.

LeBlanc et al. (2001) provided an example of the use of this mathematical model and probe procedure while programming delays to reinforcement during functional communication training. Sixteen steps were established that generally represented a 33 to 50% increase from the previous delay interval. The first step represented a 1-s delay and the final delay equaled the length of the entire session length of 10 min. The criterion for increasing the interval was a 90% or greater reduction from baseline for two consecutive sessions. They used a delay reduction (return to previous successful step) if problem behavior did not remain at a 90% reduction for two consecutive sessions. After three consecutive delay steps were successfully implemented (90% or greater reduction from baseline), a probe session three steps higher was conducted to determine if the delay increase could be accelerated. If the probe session was successful, the next probe session (3 additional steps higher) was conducted. If the probe session was unsuccessful, the previous successful level was repeated and the intermediate steps were implemented. This probe procedure allowed rapid progression of schedule thinning. In one condition, the schedule progressed from step 4 to step 14 in only 6 sessions (sessions 46-51) and in another condition from step 1 (no delay) to step 10 in the span of only 4 sessions (sessions 77-80).

## Conclusions and Future Directions

Although the model just described has worked rather well for us, it has not been systematically compared to other models. Research is needed in which this model and alternative strategies for reducing the intensity of interventions are directly compared. Specifically, researchers might directly compare schedule thinning with and without the probe procedure to determine if there is additional utility to using probes. Researchers might directly compare the preset values and probes to alternative method such as the shifting mean IRT method illustrated by Kahng et al. (2000). Researchers should also directly evaluate whether failure at a given schedule value should result in a return to a previous value or remain constant. Studies might also examine whether certain participant or client characteristics should impact our decision-making with regard to speed of schedule thinning. There is no support in the literature for differential decision-making but future studies might examine relevant characteristics such as previous history of reinforcement or dimensions of target behavior (e.g., severity, frequency, intensity). Perhaps the most important area for evaluation is the social validity of these procedures. Research should attempt to determine whether clinicians find it acceptable and easy to use values from published tables (such as those provided here) rather than directly computing values based on individuals client performance. In addition, consumers of this model will determine whether the decision criteria and recommendations prove useful in clinical or research practice and allow them to proceed through schedule thinning more rapidly than they have in the past.

Research in analogue settings, such as that conducted by Hanley et al. (2001), may be of initial value in this regard, but studies conducted in participants' natural environment ultimately will be required. Although studies have revealed a great deal about how to reduce problem behaviors using reinforcement-based procedures, most of what we know about how to reduce the intensity of such procedures is based on practical experience, not controlled research. Our current lack of empirical knowledge raises many interesting and potentially important experimental questions. For example, does

increasing the magnitude of reinforcement as the inter-reinforcer interval or ratio increases, as illustrated by Piazza et al. (1997), allow for more rapid schedule thinning? If so, what is the best way to increase magnitude? Additionally, research might attempt to determine optimal increment levels under particular conditions. The model suggested above has worked for us, but other increment values might work just as well and result in more rapid thinning. Other general areas worth investigating include strategies for determining terminal schedule values, a topic largely ignored in the literature, and techniques for incorporating rule-governed participant behavior as part of schedule-thinning procedures. Attempts to make use of findings from basic laboratory research in devising procedures for reducing the intensity of interventions is also warranted (e.g., Strohmer, McComas, & Rehfeldt, 2000). In summary, we have attempted to meet two objectives with this paper. If the present manuscript fosters research in any of these areas, we will be delighted. If it is of practical use to behavior analytic clinicians, we will be even happier because both of our objectives will be met.

## REFERENCES

- Carr, J. E., Coriaty, S., & Dozier, C. (2000). Current issues in the function-based treatment of aberrant behavior in individuals with developmental disabilities. In J. Austin & J. E. Carr (Eds), *Handbook of applied behavior analysis* (p. 91-112). Reno, NV: Context Press.
- Carr, J. E., Coriaty, S., Wilder, D. A., Gaunt, B. T., Dozier, C. L., Britton, L. N., Avina, C., & Reed, C. L. (2000). A review of "noncontingent" reinforcement as treatment for the aberrant behavior of individuals with developmental disabilities. *Research in Developmental Disabilities, 21*, 377-391.
- Carr, E. G., Yarbrough, S. C., & Langdon, N. A. (1997). Effects of idiosyncratic stimulus variables on functional analysis outcomes. *Journal of Applied Behavior Analysis, 30*, 673-686.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (1987). *Applied behavior analysis*. Columbus, OH: Merrill Publishing Co.
- Fisher, W. W., Thompson, R. H., Hagopian, L. P., Bowman, L. G., & Krug, A. (2000). Facilitating tolerance of delayed reinforcement during functional communication training. *Behavior Modification, 24*, 3-29.
- Hagopian, L. P., Fisher, W. W., & Legacy, S. M. (1994). Schedule effects of noncontingent reinforcement on attention-maintained destructive behavior in identical quadruplets. *Journal of Applied Behavior Analysis, 27*, 317-325.
- Hagopian, L. P., Fisher, W. W., Sullivan, M. T., Acquisto, J. & LeBlanc, L. A. (1998). Effectiveness of functional communication training with and without extinction and punishment: A summary of 21 inpatient cases. *Journal of Applied Behavior Analysis, 31*, 211-235.
- Hanley, G. P., Iwata, B. A., & Thompson, R. H. (2001). Reinforcement schedule thinning following treatment with functional communication training. *Journal of Applied Behavior Analysis, 34*, 17-38.
- Kahng, S. W., Iwata, B. A., DeLeon, I. G., & Wallace, M. D. (2000). A comparison of procedures for programming noncontingent reinforcement schedules. *Journal of Applied Behavior Analysis, 33*, 223-231.
- Lalli, J. S., Casey, S. D., & Kates, K. (1997). Noncontingent reinforcement as treatment for severe problem behavior: Some procedural variations. *Journal of Applied Behavior Analysis, 30*, 127-137.
- LeBlanc, L. A., Hagopian, L. P., & Maglieri, K. A. (2000). Use of a token economy to eliminate excessive inappropriate social behavior in an adult with developmental disabilities. *Behavioral Interventions, 15*, 135-143.
- LeBlanc, L. A., Hagopian, L. P., Marhefka, J. M., & Wilke, A. (2001). Effects of therapist gender and type of attention on assessment and treatment of attention-maintained destructive behavior. *Behavioral Interventions, 16*, 39-57.
- LeBlanc, L. A., Le, L., & Carpenter, M. H. (2000). Behavioral treatment. In M. Hersen and R. T. Ammerman, (Eds), *Advanced abnormal child psychology* (2nd ed.), pp. 197-218. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Logue, A. W. (1988). Research on self-control: An integrating framework. *Behavioral and Brain Sciences, 11*, 665-709.
- Martin, G. & Pear, J. (1999). *Behavior modification: What it is and how to do it* (6th ed.). Upper Saddle River, NJ: Prentice-Hall, Inc.
- Miltenberger, R. (1997). *Behavior modification: Principles and procedures*. Pacific Grove, CA: Brooks/Cole Publishing.
- Piazza, C. C., Fisher, W. W., Hanley, G. P., Remick, M. L., Contrucci, S. A., & Aitken, T. L. (1997). The use of positive and negative reinforcement in the treatment of escape-maintained destructive behavior. *Journal of Applied Behavior Analysis, 30*, 313-326.
- Piazza, C. C., Moes, D. R., & Fisher, W. W. (1996). Differential reinforcement of alternative behavior and demand fading in the treatment of escape maintained destructive behavior. *Journal of Applied Behavior Analysis, 29*, 569-572.
- Poling, A. D., & Ryan, C. (1982). Differential-reinforcement-of-other-behavior schedules: Therapeutic applications. *Behavior Modification, 6*, 3-21.
- Stokes, T. F. & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10*, 349-367.
- Strohmer, R., McComas, J. J., & Rehfeldt, R. A. (2000). Designing interventions that include delayed reinforcement: Implications of recent laboratory research. *Journal of Applied Behavior Analysis, 33*, 359-372.
- Vollmer, T. R., Iwata, B. A., Zarcone, J. R., Smith, R. G., & Mazaleski, J. L. (1993). The role of attention in the treatment of attention-maintained self-injurious behavior: Noncontingent reinforcement and differential reinforcement of other behavior. *Journal of Applied Behavior Analysis, 26*, 9-21.

Watson, T. S., & Gresham, F. M. (1998). *Handbook of child behavior therapy*. New York: Plenum.

**Author Note:** Send correspondence to Linda A. LeBlanc, Ph.D., Department of Psychology, Western Michigan University,

Kalamazoo MI 49008-5052. E-mail: [Linda.LebLANc@wmich.edu](mailto:Linda.LebLANc@wmich.edu) Portions of the paper were presented at the California Association for Behavior Analysis. We thank Jim Carr and Michelle Wallace for helpful comments on previous versions of the manuscript.



Table 1

Sample steps for schedule thinning using DRO intervals or ratio-based schedules.

<u>Step</u>	<u>Interval/Value</u>	<u>% Increase</u>
Initial Intensity: 15 s DRO		
1	30 sec	100%
2	45 sec	50%
3	60 sec	33%
4	90 sec	50%
5	120 sec	33%
6	160 sec	33%
7	230 sec	44%
8	330 sec	43%
9	450 sec	37%
10	600 sec	33%
INITIAL INTENSITY: FR 1		
1	FR2	100%*
2	FR4	100%
3	FR6	50%
4	FR9	50%
5	FR12	33%
6	FR16	33%
7	FR21	33%
8	FR28	33%

\* No smaller increment is possible.

Table 2

Hypothetical steps and probe progression for a delay to reinforcement schedule thinning procedure with an initial 1-s delay.

- 
1. After three successful steps, probe three steps higher.
  2. If probe is successful, continue to probe three steps higher
  3. If prove is unsuccessful, return to the previous successful step

Established Values		Progression of Thinning	
Step	Delay in Seconds	Completed steps	Probes
1	2	1	
2	3	2	
3	5	3	6 (success), 9 (fail)
4	7		
5	10		
6	14	6	
7	22	7	
8	35	8	
9	55	9	12 (success), 14 (success)
10	90		
11	150		
12	300		
13	450		
14	600		

---

## ON THE IDENTITY CRISIS IN OBM

Cloyd Hyten  
University of North Texas

Some confusion exists at present in the OBM community regarding the subject matter of the discipline and how concepts from outside the tradition of behavior analysis should be handled. Some of the confusion is traced to differences in historic definitions of the field, as well as issues concerning the compatibility of systems analytic concepts. Systems models used in performance-oriented analyses of organizational functioning are reviewed and compared to behavior analytic approaches. A strategy of analyzing the behavior of managers in large corporate systems is suggested as a way to better integrate systems analytic and behavior analytic approaches to understanding and improving performance.

---

Recent articles in the organizational behavior management (OBM) literature have raised questions regarding the content and direction of the area (e.g., Ghezzi, 2001; Hayes, 1999; Mawhinney, 2000, 2001). The issues revolve around the integration of concepts and models that are not behavior analytic in origin (e.g., systems models) or that appear to place behavior analysis in a secondary role. Including or excluding nonbehavioral concepts in the common language of the discipline is a practice that will have a major impact on the definition of OBM. Ghezzi, for example, observed that OBM is moving toward an identity apart from Behavior Analysis. Hayes wondered whether there is “room at the table” for nonbehavioral models in OBM. Mawhinney (2000) felt that the culture of the OBM community may be overwhelmed by the larger body of nonbehavioral models if the gates were opened to such concepts. These issues merit further discussion. In this article, I will review some of the historical sources of the “identity crisis” and address some ways to deal with it in a constructive manner.

### SOURCES OF THE IDENTITY CRISIS

Although the beginnings of OBM reach back into the 1960s, OBM began to establish a formal identity in the 1970s with the publication of numerous books as well as the founding of the *Journal of Organizational Behavior Management* in 1977 (see Dickinson, 2000, for a review of the early history of OBM). It is clear that many key people involved in this period of the field’s development saw OBM as an extension of applied behavior analysis. In other words, OBM would involve the application of behavior analytic principles to produce changes in behavior in the workplace. As Mawhinney

(2000) pointed out, the editorial that defined the mission of *JOBM* (Daniels, 1977) explicitly modeled the mission statement of Baer, Wolf and Risley (1968) in defining applied behavior analysis. It appeared that OBM would be simply behavior analysis in business and industry.

At nearly the same time, Gilbert (1978) published *Human Competence: Engineering Worthy Performance* (republished with a new foreword in 1996). Tom Gilbert had roots in behavior analysis, going back to a brief stint studying at Harvard with B.F. Skinner. Gilbert had gone on to become a pioneer in the behavioral instruction movement in the 1960s. But in *Human Competence*, Gilbert put forth a conception of a field that appeared similar to OBM in some aspects, but different in other critical aspects. He argued that the focus should be improving *performance*, where performance encompassed both behavior and worthy accomplishments. Behavior was seen as a cost to be minimized, and accomplishments as the valuable results that would enable organizations to succeed in their endeavors to bring products and services to customers. Although a field dedicated to organizational performance improvement would necessarily *involve* behavior and behavior change, it would not be *about* behavior or behavior change. For example, Gilbert pointed out that accomplishments could be improved if defective tools or equipment were replaced; some behavior change might occur with the new tools, but performance improvement was not brought about by focusing on the behavior and its associated contingencies. In fact, Gilbert (1978/1996) was quite critical of the science of behavior, and B.F. Skinner in particular, in making claims that behavior analysis (referred to

as “behavior modification” then by Gilbert) provided an adequate conceptual or practical groundwork for a discipline dedicated to performance improvement. What was truly needed in Gilbert’s opinion was not a science of behavior, but a science of accomplishment, together with a technology of how to engineer it.

Gilbert’s arguments for an accomplishment-oriented discipline represented a subtle but important difference in purpose and methods from that of OBM as it was being formulated by other behavior analysts (Gilbert referred to himself as a “behaviorist”) in the 1970s. Clearly, Gilbert did not see the discipline he was advocating as just a branch of applied behavior analysis. In fact, the logic of his emphasis on results left the door open to ideas from any discipline that could help solve a performance problem. I do not believe that this latent interdisciplinary logic was fully recognized by behavior analysts at the time. That would only become apparent in later years, as discussed below. Nonetheless, Gilbert made extensive use of behavioral concepts as part of his approach to analyzing and improving performance, most clearly evident in his widely adopted Behavior Engineering Model, which was an elaborated kind of ABC model identifying six categories of variables affecting the individual performer. His ideas influenced so many behavior analysts that he is regarded as one of the key contributors to the development of OBM (Dickinson, 2000). From the beginning, then, there have been differing conceptions of the disciplinary nature of OBM.

Gilbert’s ideas have influenced other communities that overlap to various degrees with the mainstream OBM community. Gilbert was one of the pioneers in what was then called the National Society for Programmed Instruction (NSPI), an organization of behaviorally oriented instructional designers in the 1960s. In the years following publication of his 1978 book, NSPI altered its identity to become first the National Society for Performance and Instruction (also NSPI) and later to the International Society for Performance Improvement (ISPI), taking to heart Gilbert’s ideas of a discipline of accomplishment. This is not to say that Gilbert was the sole influence on

ISPI; many other important conceptual contributions were made by people such as Dale Brethower, Robert Mager, Joe Harless, and Geary Rummler to name but a few (see Dean & Ripley, 1997). But the influence of Gilbert is undeniable. ISPI now describes its approach as human performance technology (HPT) and incorporates Gilbert’s (1978/1996) Behavior Engineering Model as the core of its cause analysis model (Dean, 1994).

The interdisciplinary logic of Gilbert’s results focus has now become manifest. Today’s HPT contains practitioners from many disciplines (see Stolovitch & Keeps, 1992; 1999), some of which (cognitivists) are hostile toward, or merely dismissive of, behavior analysis as a primitive phase they had to go through to get to a more enlightened state of understanding. In fact, Binder (1995) has lamented that behavior analysis is now a neglected minority viewpoint in HPT. Gilbert’s approach has been successful in attracting large numbers of followers. With over 10,000 members, ISPI is roughly 5 times larger than the Association for Behavior Analysis, of which OBM is only a portion. In addition, the Association for Training and Development (ASTD), an even larger professional group consisting mainly of people with an interest in corporate training, has also adopted a performance orientation and even offers a certificate in Human Performance Improvement (see [http://www.astd.org/virtual\\_community/hpi\\_courses](http://www.astd.org/virtual_community/hpi_courses)).

It may now be said that there is a large, eclectic community of performance technologists and a much smaller community of behavior analytic OBMers. The goals of each community are different. One (OBM) seeks to extend and develop behavior analytic tools for business and industry; the other (HPT) seeks effective technologies regardless of theoretical origin for solving performance problems. There is some overlap in the membership of both communities (I am a member of both, for example) as well as intellectual overlap. It is the overlap that is causing the identity crisis in OBM. The critical question becomes: should OBM maintain its distinct behavior analytic

identity, or should it incorporate useful concepts and models from the broader performance technology community? Were there no useful ideas in the HPT community there would be no intellectual basis for an identity crisis in OBM; the only remaining basis would be the attraction of belonging to a larger community that has achieved more public acceptance than the small OBM community. The majority of the intellectual overlap at present is due to the influence of general systems theory and the resulting systems analytic models. These conceptions have been the focus of many of the questions raised regarding the theoretical identity of OBM.

### Systems Models and OBM

Systems concepts have permeated disciplines ranging from physiology to ecology. With their emphasis on interacting parts making up the whole, systems approaches would seem an obvious fit to the analysis of organizational functioning. Indeed, organizational pundits such as Senge (1990) have deemed systems concepts one of the essential disciplines relevant to business management. Brethower (1972; 1982; 1995; 2001) was an early and persistent advocate applying systems concepts to help analyze and solve performance problems in organizations. He devised a generic, scaleable model of an entity (e.g., a company, or a department within that company) viewed as a system and referred to it as the Total Performance System (TPS; see Fig.1). The TPS describes critical elements such as inputs, processes, outputs, customers (the “receiving system”) and regulating feedback. As a model, the TPS is useful in providing the “big picture” of what comprises organizational functioning.

Rummler and Brache (1995) devised a more elaborate systems model (the super-system or adaptive system model; see Fig. 2) that includes additional elements such as competitors and other environmental influences affecting how an organization functions. In addition, Rummler and Brache proposed 3 levels of variables that drive performance: *organization level* variables such as company goals, *process level* variables such as the design of workflow through the organization, and *performer level*

variables that encompass the elements of Gilbert’s Behavior Engineering Model affecting individuals in the performance of their job duties. Although performer level issues are included in the Rummler-Brache model, the weight of their analysis focuses on the previously neglected process level. Their analysis of process problems spawned a huge interest in process mapping (drawing graphical representations of workflow) and process re-design as a strategy to improve organizational functioning (see, e.g., Malott, 2001).

These systems models are clearly compatible with a general performance improvement orientation, but they raise questions about their compatibility with the traditional principles of behavior analysis and with the goals of OBM. That has made some behavior analysts (e.g., Ghezzi, 2001; Hayes, 1999; Mawhinney, 2000) uncomfortable. What is it about systems models that seem to bother some behavior analysts? There are several answers. Systems models focus on the organization as a whole and the results the organization must produce to stay in business. They are fundamentally conceptions about organizations, not about behavior per se. Issues focusing on behavior are mentioned in systems models, but they are often given a status secondary to other systems issues such as process inefficiencies. Behavior seems to disappear in the models. This may be somewhat threatening to behavior analysts as their discipline is marginalized to some degree in systems approaches.

Systems operate according to principles other than those that govern behavior. Hayes (1999) made just this point, “Systems are not maintained by reinforcement. They aren’t extinguished by its absence, and they don’t come under stimulus control.” (p. 65). Hayes wondered further whether behavior analysis and systems analysis are operating from the same philosophical foundations. Systems models include concepts that are philosophically or theoretically troublesome for behavior analysts. Performance is often described as goal-directed and regulated by feedback (see Brethower, 2001), concepts that behavior analysts have wrestled with for some time. The use of the

term feedback alone has generated more than its fair share of debate in OBM over the years (e.g., Duncan & Bruwelheide, 1986; Normand, Bucklin, & Austin, 1999; Peterson, 1982).

On the other hand, there are some similarities to behavior analysis in systems approaches that have attracted the interest of the behavior analytic community for decades. Willems (1974) brought systems approaches to the attention of applied behavior analysts in his discussion of behavioral ecology. He pointed out that phenomena such as unintended side effects of behavioral interventions could be understood to be systems effects. Krapfl and Gasparatto (1982) saw systems analysis and behavior analysis as compatible approaches, capable of being blended into “behavioral systems analysis” for use in OBM. They pointed out that the core systems concepts of input, processes, and outputs together with feedback parallel the familiar behavioral concepts of antecedents, behavior, and consequences. Recently, Redmon and Mason (2001) have incorporated the systems models of Brethower and Rummel into a coherent framework for understanding corporate cultures. They proposed that their systems framework could be used to change cultures and provided some examples supporting that notion.

There is another feature of these systems models that enables them to be more closely compared to behavior analytic concepts: organizational systems are described as adapting to their environments in response to the actions of the marketplace. Rummel and Brache (1995) put it succinctly, “A processing system (organization) will either adapt to its environment, especially its receiving system (market), or cease to exist.” (p. 12). Mawhinney (2000; 2001) realized that this adaptive property of organizational systems models is similar to the modern behavior analytic emphasis on *selection* as a causal process exemplified in Glenn’s (1988; 1991) concept of the metacontingency. Thus, there is some common ground between systems analysis and behavior analysis.

Glenn’s (1988) formulation of the metacontingency was intended to address broad

issues of cultural practices and the mechanisms of their selection by consequences. Within a short time, organizational behavior analysts applied this concept to the study of organizational practices and organizational survival (Mawhinney, 1992; Redmon & Agnew, 1991; Redmon & Wilk, 1991). As applied to organizations, a metacontingency refers to the relation between organizational practices and those outcomes critical to the survival and success of the organization (such as profitability). The metacontingency selects contingencies within the organization that lead to effective organizational practices (those that produce organizational outcomes that meet the survival requirements of the ambient business environment). Metacontingencies of business survival also exert selective pressure against organizational contingencies and their associated practices that are not well aligned with the demands of the marketplace and business competitors. Such organizational cultures will perish. The concept of the metacontingency gives behavior analysts a tool to discuss macro-level phenomena such as organizational adaptation as instances of a common mechanism of selection- the same mechanism that operates at the level of the individual in the evolution of operant behavior and at the level of the species in biological evolution.

### **Can Systems Analysis and Behavior Analysis be Integrated?**

Does the recognition of similarities between metacontingencies and adaptive organizational systems models like Brethower’s TPS or Rummel and Brache’s Super-system mean that systems approaches and behavior analytic approaches address the same issues with different language? On the surface it would seem so. One can make a case that the organizational systems models are fleshed out graphical representations of metacontingencies. In this sense, systems models outline some of the key elements involved in metacontingency selection processes. This is a step in the direction of bringing systems analysis and behavior analysis closer together, but I don’t think it is sufficient to make behavior analysts fully comfortable with systems approaches. What is needed is something that would

highlight the role of behavior within the larger system. There is just such an issue at the core of the selection/adaptation processes identified by the metacontingency and systems models. It is not the rather obvious and readily understood behavior of the performers carrying out work activities in the subsystems comprising the larger processing system. Rather, it has to do with the function labeled *management* within the processing system (refer to Fig. 2).

Rummler (in Rummler, 2001; Rummler & Brache, 1995) has been quite explicit about the role managers play in adjusting organizational practices to adapt to organizational survival requirements. His discussions of management practices make it clear that layers of management personnel monitor and adjust various components of the organization, from the strategic to the tactical level, in an effort to align the organization with the requirements of the marketplace. Redmon and Mason (2001) made a similar observation, "One responsibility of management involves monitoring external demands and setting conditions within an organization to support adaptive practices." (p. 439). This element is missing from the metacontingency concept, which, in striving to be generically applicable to any cultural entity, leaves open the details or mechanics of how lower-level contingencies and practices are developed, eventually to be selected by metacontingencies of organizational survival. Business organizations may be somewhat of a special case in that, unlike other cultural entities such as society at large, most organizational practices and contingencies are deliberately planned, designed, and adjusted by a cadre of management personnel. So, it is the systems model that points to a crucial behavioral issue, in principle analyzable by behavior analysts, but so far little analyzed- the behavior of managers in making adjustments to organizational policies and practices that affect organizational success and survival.

The behavior of those who design and adjust organizational practices is the missing behavioral linkage between systems level analyses and behavior analyses. Although Hayes' (1999) statement that systems are not reinforced or extinguished is true, the behavior

of those who design and modify them undoubtedly is affected by consequences. A behavior analytic account of the variables affecting this management behavior would bring behavior to the forefront within organizational systems theory. This would go a long way toward solving the identity crisis in OBM brought on by systems models in which, at present, behavior seems to disappear.

OBM and systems analytic consultants already intervene in organizations through changing the behavior of management personnel. The consultant may train managers in the finer aspects of reinforcement delivery to their subordinates (see. e.g., Braksick, 2000) or hold multi-day workshops in which managers learn to revise defective cross-functional processes. Managers are seen as tools in producing performer-level changes, where the performer is an incumbent in a lower-level job creating the products or services of the firm. OBM has developed some proficiency in modifying such performer-level issues through the actions of managers. But this involves engineering the behavior of managers, and it is not the same as understanding the variables that influence it when a consultant is not directing or guiding them. More attention needs to be devoted to the analysis of the manager as a performer when he or she seeks and interprets receiving system feedback, makes decisions regarding company policies that will affect critical outcomes, or plans for future organizational growth. These are just some of the managerial behaviors that comprise the necessary actions involved in what Mawhinney (2001) referred to as "minding the metacontingencies."

#### **A Real Challenge for OBM**

Analyzing managerial decision-making and planning activities will be no easy task for behavior analysts. Such complex human behavior likely has certain characteristics that play to the weaknesses in current behavior analytic theory, not the strengths. I will speculate here on some of these characteristics. Much decision-making behavior in organizations is probably either covert on the part of executives or occurs as part of a social

interaction in a committee. Critical strategic level decisions are bound to occur with very low frequency. They are also likely influenced by long personal histories acting in concert with rules derived from business lore or from advice given by business colleagues or management gurus. The consequences of decisions or plans made are likely to involve a multitude of events, many of them temporally remote (on the order of months and years) and distributed across groups of people. In addition, the very business environment to which managers are striving to adapt is highly dynamic, with competitors and customers redefining what is a valued product or service in relatively short time spans. This is not the profile of work activities so commonly addressed by current OBM methods: the higher frequency overt daily work activities of salespeople, waitstaff, call center operators, manufacturing line workers, and so on (see Nolan, Jarema, & Austin, 1999, for an analysis of publication content in *JOBM*).

Does OBM have the conceptual tools needed for the analysis of managerial behavior? At present, useful behavior analytic accounts of something like executive decision-making or planning simply do not exist. Concepts that might be applied to address the issues include rule-governed behavior and indirect-acting contingencies, but there is little consensus on their definition or utility at least on the part of OBM writers (cf. Agnew & Redmon, 1992; Malott, 1992; Mawhinney, 2001). Even a leading OBM figure like Hopkins (1999), for example, finds rule-governed behavior an empty concept and rejects it as belonging to the corpus of mainstream behavior analytic theory. Normand, Bucklin, & Austin (1999), on the other hand, argue that an understanding of rule-governed behavior is critical for OBM analyses.

Addressing low frequency behaviors may be very difficult. Behavior analysis deals with higher frequency behaviors much more readily than very low frequency behaviors for several reasons. First, it is easier to detect causal relationships with increasing repetition, so higher frequency behavior is more readily analyzed. Second, from an applied standpoint, a higher frequency of behavior provides more opportunities for contact with consequences.

Explaining the social dynamics of behavior in committees would also present an obstacle. Despite the noble efforts of behavior analysts like Guerin (1994), any realistic assessment would have to conclude that behavior analytic conceptions of social behavior are still in a somewhat underdeveloped stage.

It is possible that new concepts may be needed for the analysis of managerial behavior, but behavior analysis has shown itself to have a rather slow cycle time from conceptual development to acceptance and use by a significant proportion of the community. Witness the concept of the establishing operation. Michael (1982) proposed it twenty years ago, and it is just within the last few years that the concept is being discussed at any length in the *Journal of Applied Behavior Analysis* (e.g., McGill, 1999; Michael, 2000; Northup, Fusilier, Swanson, Roane, & Borrero, 1997; Smith & Iwata, 1997) or in the OBM literature (see Agnew, 1998; Hyten, 2001). It is notable that it has yet to receive any experimental treatment in the pages of the field's premier basic research journal, the *Journal of the Experimental Analysis of Behavior*.

Analyzing the behavior of managers within large systems is likely to be a long-range project for OBM. In the meantime, behavior analysts should feel a bit more comfortable using current systems models. There are important, complex, and fascinating behavior issues embedded within those box diagrams if we will only look for them. Systems concepts have proven that they are applicable to many different disciplines, and the work of Brethower and Rummel have shown that systems models are a good fit to the analysis of organizational performance. Systems models and their advocates are certainly not anti-behavioral; at the very worst they are neutral with respect to the role of behavior. In fact, systems advocates like Brethower have been among those reminding the broader performance technology community of the benefits of the behavioral approach (see Brethower, 2000). It would be foolish to shun such allies. In my opinion, the OBM community should accept current systems analytic models as valuable tools possessing a reasonable degree of compatibility with behavior



analytic approaches while, at the same time, strive to develop an understanding of the behavior involved in the components of the system. Eventually, this may lead to a grand synthesis of the two approaches and there will be a unified model truly worthy of the name "behavioral systems analysis." Perhaps then the identity crisis in OBM will be resolved.

## REFERENCES

- Agnew, J. L. (1998). The establishing operation in organizational behavior management. *Journal of Organizational Behavior Management, 18*(1), 7-19.
- Agnew, J. L., & Redmon, W. K. (1992). Contingency specifying stimuli: The role of "rules" in organizational behavior management. *Journal of Organizational Behavior Management, 12*(2), 67-76.
- Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1*, 91-97.
- Binder, C. (1995). Promoting HPT Innovation: A return to our natural science roots. *Performance Improvement Quarterly, 8*(2), 95-113.
- Braksick, L. W. (2000). *Unlock behavior: Unleash profits*. New York: MacGraw-Hill.
- Brethower, D. M. (1972). *Behavioral analysis in business and industry*. Kalamazoo, MI: Behaviordelia.
- Brethower, D. M. (1982). The total performance system. In R. O'Brien, A. Dickinson, & M. Rosow (Eds.), *Industrial behavior modification* (pp. 250-369). New York: Pergamon.
- Brethower, D. M. (1995). Specifying a human performance technology knowledge base. *Performance Improvement Quarterly, 8*(2), 17-39.
- Brethower, D. M. (2000). Integrating theory, research, and practice in human performance technology: Examples from behavioral, cognitive, and constructivist theories. *Performance Improvement, 39*(4), 33-43.
- Brethower, D. M. (2001). Managing a person as a system. In L. J. Hayes, J. Austin, R. Houmanfar, & M. C. Clayton (Eds.), *Organizational change* (pp. 89-105). Reno, NV: Context.
- Daniels, A. C. (1977). Editorial. *Journal of Organizational Behavior Management, 1*(1), v-vii
- Dean, P. J. (Ed.). (1994). *Performance engineering at work*. Batavia, IL: International Board of Standards for Training, Performance and Instruction.
- Dean, P. J., & Ripley, D. E. (Eds.). (1997). *Performance improvement pathfinders*. Washington, DC: ISPI.
- Dickinson, A. M. (2000). The historical roots of organizational behavior management in the private sector: The 1950s-1980s. *Journal of Organizational Behavior Management, 20*(3/4), 9-58.
- Duncan, P. K., & Bruwelheide, L. R. (1986). Feedback: Use and possible behavioral functions. *Journal of Organizational Behavior Management, 7*(3/4), 91-114.
- Ghezzi, P. M. (2001). Discussion of D.M. Brethower. In L. J. Hayes, J. Austin, R. Houmanfar, & M. C. Clayton (Eds.), *Organizational change* (pp. 107-109). Reno, NV: Context.
- Gilbert, T. F. (1978/1996). *Human competence: Engineering worthy performance*. Washington, DC: International Society for Performance Improvement and HRD Press.
- Glenn, S. S. (1988). Contingencies and metacontingencies: Toward a synthesis of behavior analysis and cultural materialism. *The Behavior Analyst, 11*, 161-179.
- Glenn, S. S. (1991). Contingencies and metacontingencies: Relations among behavioral, cultural, and biological evolution. In P. A. Lamal (Ed.) *Behavioral analysis of societies and cultural practices* (pp. 39-73). Washington, DC: Hemisphere.
- Guerin, B. (1994). *Analyzing social behavior: Behavior analysis and the social sciences*. Reno, NV: Context.
- Hayes, L. J. (1999). Dining with the Devil: A response to Normand, Bucklin and Austin. *Journal of Organizational Behavior Management, 19*(3), 63-65.
- Hopkins, B. L. (1999). The principles of behavior as an empirical theory and the usefulness of that theory in addressing practical problems. *Journal of Organizational Behavior Management, 19*(3), 67-74.
- Hyten, C. (2001). Theoretical issues in the design of self-managed work teams. In L. J. Hayes, J. Austin, R. Houmanfar, & M. C. Clayton (Eds.), *Organizational change* (pp. 325-344). Reno, NV: Context.
- Krapfl, J., & Gasparatto, G. (1982). Behavioral systems analysis. In L. W. Frederiksen (Ed.), *Handbook of organizational behavior management* (pp. 21-38). New York: Wiley.
- Malott, M. E. (2001). Putting the horse before the cart: Process-driven change. In L. J. Hayes, J. Austin, R. Houmanfar, & M. C. Clayton (Eds.), *Organizational change* (pp. 297-320). Reno, NV: Context.
- Malott, R. W. (1992). A theory of rule-governed behavior and organizational behavior management. *Journal of Organizational Behavior Management, 12*(2), 45-65.
- Mawhinney T. C. (1992). Evolution of organizational cultures as selection by consequences: The Gaia hypothesis, metacontingencies, and organizational ecology. *Journal of Organizational Behavior Management, 12*(2), 1-26.
- Mawhinney, T. C. (2000). OBM today and tomorrow: Then and now. *Journal of Organizational Behavior Management, 20*(3/4), 73-137.
- Mawhinney, T. C. (2001). Organization-environment systems as OBM intervention context: Minding your metacontingencies. In L. J. Hayes, J. Austin, R. Houmanfar, & M. C. Clayton (Eds.), *Organizational change* (pp. 137-165). Reno, NV: Context.
- McGill, P. (1999). Establishing operations: Implications for the assessment, treatment, and prevention of problem behavior. *Journal of Applied Behavior Analysis, 32*, 393-418.
- Michael, J. (1982). Distinguishing between discriminative and motivational functions of stimuli. *Journal of the Experimental Analysis of Behavior, 37*, 149-155.
- Michael, J. (2000). Implications and refinements of the establishing operation concept. *Journal of Applied Behavior Analysis, 33*, 401-410.

- Nolan, T., Jarema, K., & Austin, J. (1999). An objective review and comparison of the *Journal of Organizational Behavior Management: 1987-1997*. *Journal of Organizational Behavior Management, 19*(3), 83-114.
- Normand, M., Bucklin, B., & Austin, J. (1999). The discussion of behavioral principles in JOBIM. *Journal of Organizational Behavior Management, 19*(3), 45-56.
- Northup, J., Fusilier, L., Swanson, V., Roane, H., & Borrero, J. (1997). An evaluation of methylphenidate as a potential establishing operation for some classroom reinforcers. *Journal of Applied Behavior Analysis, 30*, 615-625.
- Peterson, N. (1982). Feedback is not a new principle of behavior. *The Behavior Analyst, 5*, 101-102.
- Redmon, W. K., & Agnew, J. L. (1991). Organizational behavioral analysis in the United States: A view from the private sector. In P. A. Lamal (Ed.) *Behavioral analysis of societies and cultural practices* (pp. 125-139). Washington, DC: Hemisphere.
- Redmon, W. K., & Mason, M. A. (2001). Organizational culture and behavioral systems analysis. In Johnson, C. M., Redmon, W. K., & Mawhinney, T. C. (Eds.) *Handbook of organizational performance: Behavior analysis and management*, (pp. 437-456). New York: Haworth.
- Redmon, W. K., & Wilk, L. A. (1991). Organizational behavioral analysis in the United States: Public sector organizations. In P. A. Lamal (Ed.) *Behavioral analysis of societies and cultural practices* (pp. 107-123). Washington, DC: Hemisphere.
- Rummmler, G. A. (2001). Performance logic: The organization performance Rosetta Stone. In L. J. Hayes, J. Austin, R. Houmanfar, & M. C. Clayton (Eds.), *Organizational change* (pp. 111-132). Reno, NV: Context.
- Rummmler, G. A. & Brache, A. P. (1995). *Improving performance: Managing the white space on the organization chart*. (2<sup>nd</sup> edition). San Francisco: Jossey-Bass.
- Senge, P. M. (1990). *The fifth discipline: The art and practice of the learning organization*. New York: Doubleday.
- Smith, R. G., & Iwata, B. A. (1997). Antecedent influences on behavior disorders. *Journal of Applied Behavior Analysis, 30*, 343-375.
- Stolovitch, H. D., & Keeps, E. J. (Eds.). (1992). *Handbook of human performance technology*. San Francisco: Jossey-Bass.
- Stolovitch, H. D., & Keeps, E. J. (Eds.). (1999). *Handbook of human performance technology* (2<sup>nd</sup> edition). San Francisco: Jossey-Bass.
- Willems, E. P. (1974). Behavioral technology and behavioral ecology. *Journal of Applied Behavior Analysis, 7*, 151-165.

**Author Note:** Address correspondence to Cloyd Hyten, Department of Behavior Analysis, P.O. Box 310919, Denton, TX, 76203. Email: [Hyten@scs.cmm.unt.edu](mailto:Hyten@scs.cmm.unt.edu)

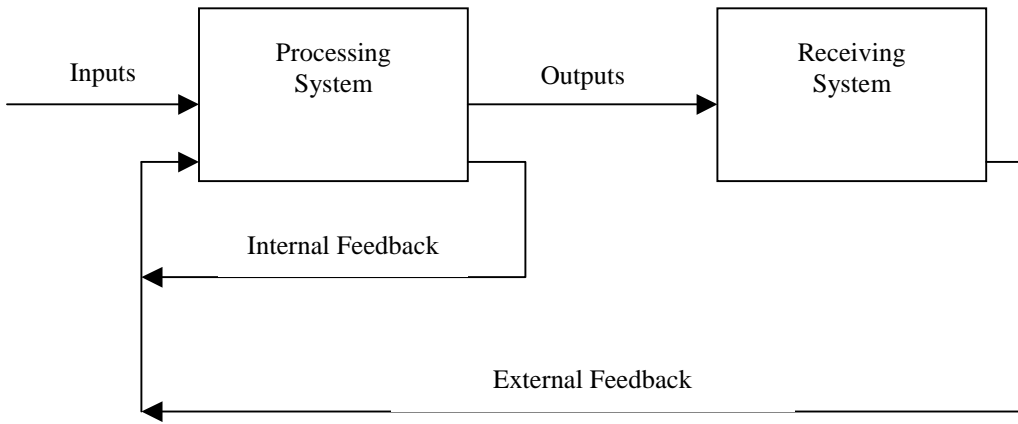


Figure 1. A basic systems model based on The Total Performance System developed by Brethower.

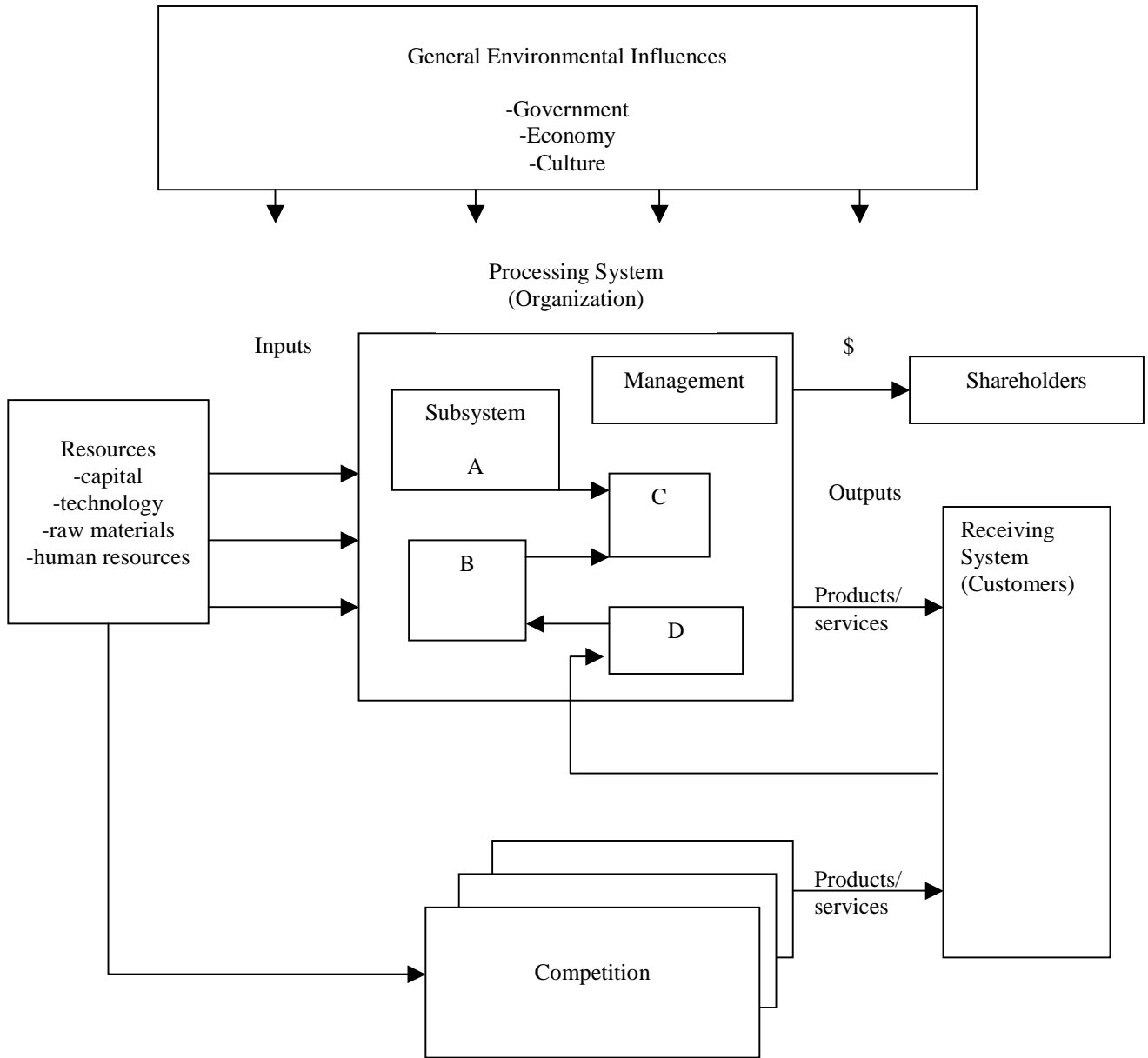


Figure 2. The Super-system developed by Rummler and Brache.

## PERFORMANCE-BASED PAY IN THE WORKPLACE: MAGIC POTION OR MALEVOLENT POISON?

Brad Gilbreath

Indiana University-Purdue University Fort Wayne

Michael M. Harris

University of Missouri-St. Louis

Performance-based pay systems have been widely used in organizations. Despite their popularity, they have been criticized on a number of grounds, such as their lack of effectiveness. There is also surprisingly little research on what effects such systems may have on employee perceptions. The present paper is divided into two sections. First, we review several of the major criticisms of performance-based pay systems and provide responses in their defense. Second, based on research literature, we offer a discussion of numerous perceptions and attitudes that may mediate the relationship between performance-based pay and employee performance.

---

"Money may kindle, but it cannot by itself, and for very long, burn." -Igor Stravinski

"Money, the root of all evil...but the cure for all sadness." -Mike Gill

"Money is the barometer of a society's virtue." -Ayn Rand

There are many opinions regarding money. Today, many organizations assume that money is positively viewed and can be an important employee motivational tool. In earlier times, work motivation was not a significant problem. If you wanted to eat, you had to work. If your harvesting and hunting efforts were unsuccessful, you went hungry. In Western industrial societies, this has changed. A social safety net exists, making the need to work less urgent than in the past. Work motivation also presents a greater challenge because many of us work as hired labor rather than as independent craftpersons or farmers. If organizations are to use money effectively as a motivator, an understanding of how financial incentives affect employees is essential.

As indicated by our title, some have insisted that performance-based pay not only has little positive effect, but also may have a negative impact on the organization. Our aim in this article is to identify and explain the variety of effects performance-based pay may have on employees. We begin by providing an overview of performance-based pay. We then discuss some of the criticisms of performance-based pay and offer some potential responses to these criticisms. Finally, we describe some possible

mediating factors that affect the relationship between performance-based pay systems and job performance.

### Overview of Performance-Based Pay

Prior to the late 1800s, a large percentage of workers were basically self-employed. However, the industrial revolution changed much of this in the U.S. during the late 1800's and continued to during the twentieth century. Large manufacturing facilities developed, employing hundreds and, eventually, thousands of workers. Motivating effective work behavior from these legions of employees was essential to the success of such enterprises. Still, however, these were simpler times. Until unions became more firmly entrenched in the 1930's, motivation could be achieved by the threat of losing one's job. The threat of unemployment was salient to workers, particularly during the recurrent depressions that bedeviled the U.S. economy. Fear *is* a motivator, and the threat of losing one job when other jobs were scarce and no unemployment benefits were available was a fearful specter.

The threat of losing one's job was not the only fear-based motivator. Prior to the influx of unions and more enlightened norms of supervision, motivation was sometimes achieved by fear of one's supervisor. In some organizations, discipline and productivity were achieved through physical coercion. Henry Ford is said to have selected some of his production foremen on the basis of their ability to intimidate workers; the bigger and scarier the foremen, the

better. According to legend, Ford selected his security chief during a plant tour when he observed this individual beating up a worker.

As time went on, unionization and other changes to the industrial setting made more sophisticated approaches to motivation necessary. The challenge was how to induce employees to perform their jobs effectively when there were too many workers for close monitoring. As work became more mechanized and monotonous, and as spans of control increased in the burgeoning factories, the use of financial incentives, such as piece-rate pay, was an easy choice for shaping employee behavior. It did not take long, however, for skeptics of the use of financial incentives to emerge. Chester Barnard (1938), for example, believed that financial incentives had taken on exaggerated importance as motivational tools, reasoning that once enough money has been earned for physiological necessities, material incentives (including money) become weak in their effect. He felt that the widely held belief in the effectiveness of material incentives was an illusion. In his words,

“it seems to me to be a matter of common experience that material rewards are ineffective beyond the subsistence level excepting to a very limited proportion of men; that most men work neither harder for more material things, nor can be induced thereby to devote more than a fraction of their possible contribution to organized effort.” (p. 144).

Barnard also believed that organizations’ lack of skill in offering non-material inducements (e.g., feelings of pride and accomplishment; satisfactory social conditions) lead them to substitute material inducements for non-material ones. He was skeptical about the effectiveness of this approach:

“Under favorable circumstances, to a limited degree, and for a limited time, this substitution may be effective. But to me, at least, it appears utterly contrary to the nature of men to be sufficiently induced by material or monetary considerations to contribute enough effort to a cooperative system to enable it to be productively efficient to the degree necessary for persistence over an extended period.” (p. 93).

Performance-based pay is nothing new; in ancient times Mesopotamians were paid by the basket for picking olives (Wiscombe, 2001). Unfortunately, although the technique has been used for quite some time, we still do not have a firm grasp on how, when, and why performance-based pay works and does not work. What Opsahl (1967) wrote 35 years ago still seems valid: “Although money is the main method by which we reward and modify behavior in industry, very little is actually known about its effects” (p. 208). Although this is an irritant to those of us who would like some straightforward guidelines for effectively designing compensation systems, it is not surprising. As Gellerman (1963) aptly expressed: “Monetary incentives become quickly entangled with a lot of other motives that have little or nothing to do with money, so that the ultimate effect of money itself is no easier to identify than is an egg in an omelet” (p. 63).

Performance-based pay is a compensation program where at least some amount of pay depends on a pre-determined level of performance. Beyond this simple definition, however, there are many different kinds of performance-based pay programs. We will focus on performance-based pay programs that are based on an objective formula for determining the size of the payout. Thus, merit pay programs are not reviewed or considered here. Performance-based pay programs (excluding merit pay programs) are often categorized into the following categories:

1. individual incentive programs (e.g., commissions);
2. team-based incentives;
3. organization/plant-based incentives (e.g., profit-sharing).

Because there are different performance-based pay programs, some may be more effective than others. In addition, different performance-based pay programs may have different effects on motivation. Most practitioners and scholars have identified both strengths and weaknesses of each type of program, and it is widely recognized that a combination of such programs is often necessary

to effectively motivate employees (e.g., Harris, 2000). One final point: a poorly designed and implemented performance-based pay program is bound to fail, while a well-designed and implemented performance-based pay program may be reasonable successful.

### Performance-Based Pay: Criticisms and Responses

Performance-based pay seems to be a subject of a surprising amount of controversy compared to many other human resource practices. And it is telling that, even after decades of experience with and research on performance-based-pay, we continue to see articles with titles such as “Can Pay for Performance Really Work?” (Wiscombe, 2001). Why is this the case? Rewarding people financially based on their contribution to the organization, on the face of it, does not seem controversial. Translating this idea into effective practice, however, has been problematic. In this section of the article we discuss some criticisms and offer our responses. Primary areas of controversy include whether performance-based pay is effective, whether performance-based pay is the best motivator, and whether it is too difficult to design and implement to be useful.

#### CRITICISM: PERFORMANCE-BASED PAY DOES NOT WORK WELL

Probably the most challenging argument to performance-based pay systems is that they are not an effective means of improving performance. We have reviewed the literature on performance-based pay, and we can tell you that it is extensive yet contradictory. Fortunately, G. Douglas Jenkins and his colleagues (Jenkins, Mitra, Gupta & Shaw, 1998) provided some order to this literature with a meta-analysis of the performance-based pay literature. Very briefly, a meta-analysis is a quantitative summary of a body of literature. This approach has been applied to a large number of topics in psychology, ranging from the effectiveness of psychotherapy to the accuracy of personnel testing. Jenkins and his colleagues located 39 studies on the effect of *individual* financial incentives on work performance. Based on these studies, they found a corrected correlation of .34 between financial incentives and work quantity; the corrected correlation between financial

incentives and work quality was lower (.08). Thus, this research suggested there is a reasonable, though modest, effect of financial incentives on work quantity, but almost no effect on quality. In addition, they compared the effect of financial incentives on performance in a laboratory setting versus a field setting and found that the latter context produced bigger effects than the former. The key conclusion is that financial incentives do have a positive effect of a reasonable magnitude on quantity; while the effect on quality is smaller, it is *not* negative. Thus, there was no indication that individual incentives hurt quality.

Despite this relatively positive conclusion, a review of the literature can lead one to conclude that performance-based pay does not have a completely positive image. In a study of state government pay-for-performance systems, only two out of 20 personnel directors ranked such programs as effective (Ingraham, 1993). More recently, only 22 percent of the companies responding to a Hewitt Associates survey said they believe pay incentives work (Wiscombe, 2001). Nevertheless, as reviewed above, the research literature does support the effectiveness of performance-based pay programs when objective measures of performance were used. This does not mean that poorly designed and implemented programs will be successful; it does indicate that performance-based pay programs can work if done properly.

#### CRITICISM: PERFORMANCE-BASED PAY IS NOT THE BEST MOTIVATOR

A corollary to the previous criticism is that, while performance-based pay may not be completely useless, it is not the *best*, or even among the best, workplace motivators. In reply to the meta-analysis described above, a critic might argue that a corrected correlation of .34 is only modest and not even close to a perfect correlation of 1.00. As a result, other workplace interventions (e.g., improving job design) would be far more effective.

We have several responses to this criticism. First, the relatively modest effect of financial incentives may be at least in part explained by a closer look at job performance. According to most theorists, job performance is

broadly determined by at least two major factors: motivation and abilities. To borrow a favorite personal example, one could offer most people \$20 million for each home run they hit against major league pitchers and the result would be same if you offered them \$5 per home run. In either case, they will fail to hit any home runs! That's because it takes both motivation *and* ability to hit home runs against major league pitching. We should therefore not dismiss the relatively modest effect of performance-based pay on performance; performance-based pay alone is not enough to make huge improvements in performance.

Second, and related to our first response, the best companies use a number of motivational tools, including performance-based pay, to motivate workers. Few experts would recommend that organizations use one motivational tool to the neglect of others; a combination of effective selection practices, pay practices, job-design techniques and leadership approaches are needed for maximum effectiveness. We certainly would not recommend that companies rely *solely* on performance-based pay to motivate the workforce.

Third, one cannot merely *assume* that other motivators are more effective. In an examination of the job design literature, for example, Fried (1991) examined the correlations between job characteristics (e.g., variety, feedback) and worker performance. He reported average raw correlations ranging from .06 to .15 between different job characteristics and a measure of job performance. Corrected correlations ranged between .09 (work variety) and .22 (feedback). Compared to the corrected correlation of .34 for financial incentives, the relationships for job characteristics, at least on an individual-dimension basis, are lower. Thus, we question the assumption that there are other motivational tools that are superior to financial incentives. Before criticizing the effect of financial incentives, then, we need empirical evidence of the superiority of other approaches.

Finally, an implicit assumption is often made that a performance-based pay system works only when employees need or desire money. As people work up to a comfortable

financial status, it is believed that one could expect that pay will begin to lose its ability to bring about highly desired outcomes for many individuals. Other outcomes, such as a satisfactory balance between home life and work life may become more important. We would like to point out, however, that performance-based pay provides more than just more money for successful employees. It can sometimes satisfy other needs, such as esteem needs, if income is regarded as a source of status. Barnard (1938) noted that income can serve as "an index of social status" (p. 145). Wang, Kick, Fraser, & Burns (1999) note that socioeconomic status can play an important role in shaping individuals' views of themselves. This shaping can happen through the effects of reflected appraisals, or how employees see themselves from the point of view of others (Wang et al.). Thus, an employee's status as perceived by others can be expected to affect his or her self perceptions. In this way, financial rewards based on effective job performance could be expected to affect employees' identity, self-worth, vocational status, and quality of life. How these self-perceptions would translate into performance, however, would be complicated, if not impossible, to predict.

Money also satisfies employees' need for achievement. One sometimes hears people talk about income as a way of keeping score. In other words, some people use money as an indicant of how well they are doing in the game of life. Money, in this sense, can serve some employees as a relatively clear-cut way of measuring personal achievement. For employees with this mind-set, performance-based pay may be an effective way to focus their need for achievement in directions helpful to the organization. Certainly other means of recognizing achievement should also be utilized in addition to financial rewards. People have other needs that are ignored when behavior is shaped only through monetary incentives. For example, needs for growth and self-actualization will be ignored, as will needs for love and relatedness (i.e., meaningful relationships with others). For more information on the meaning of money, see Mitchell (1999).

Nevertheless, we agree that pay is not necessarily the most effective motivator for all



employees. Most pay systems function in a manner similar to the pre-cafeteria benefits programs, where all employees received the same benefits mix. Eventually someone had the smart idea of allowing employees to pick and choose their benefits mix to meet their specific needs. Reward systems could profit from similar modification. We know that employees are coming to value organizational rewards other than pay. "Give-me-a-life" benefits (e.g., flexible schedules; more time off from work) seem to be favored by many employees presently. Similarly, organizations might get more motivating power from a mix of performance-based rewards. For instance, some employees could opt for additional time off from work rather than a performance bonus. Some employees might want their reward for effective job performance, not in the form of a pay increase, but as a donation to a charitable cause of their choice or as additional vacation time. It may be worthwhile to permit employees more latitude in terms of using pay incentives to pursue personally desired outcomes. A key challenge, then, would be restructuring the performance-reward system to accommodate a greater variety of rewards. Another challenge would be identifying individual employee's primary motivators.

### **Criticism: Designing And Developing Effective Performance-Based Pay Systems Is Too Difficult**

To be effective, performance-based pay systems require considerable design and planning. As Flannery, Hofrichter, and Platten (1996) point out, performance-based pay programs must be supported by highly effective performance management and employee communication programs. And both performance management and employee communications are perennially problematic issues for organizations. Performance appraisal may be second only to performance-based pay in terms of the level of controversy about its problems and lack of effectiveness. Therefore, in many organizations, performance-based pay's foundation--the performance appraisal system--will be less than solid. Until issues related to performance management and measurement, as well as employee communication and understanding have been resolved, performance-based pay seems like a questionable proposition.

We have two responses to this criticism. First, we believe that any good human resource management system requires considerable design and planning effort, particularly when such an important issue is at stake. Moreover, to state that such related systems, such as performance appraisal, are often not well designed and implemented is simply making an excuse for why a solid performance-based pay system is not in place.

Second, many performance-based pay systems use highly objective pay programs that avoid reliance on subjective performance appraisal systems. At Nucor Steel, which has experienced quite a bit of success with its compensation system, measurement and communication of results is straightforward. As Collins (1999) described the system: "Nucor has no discretionary bonuses. It's more like a sports bonus system: if you score so many points or win a certain number of races, you get a bonus based on a predetermined formula. Period." (p. 75). The key then is that any good human resource management system takes effort and time and knowledgeable experts to design and implement. In short, we do not believe that performance-based pay systems are impossibly difficult to design and implement. They do, however, require time and effort; but, if done properly, it should be time and effort well spent. While performance-based pay systems may not be appropriate for every job and every organization, many organizations would benefit from using them.

Next we describe some of the variables that may explain the relationship between performance-based pay and job performance. In other words, what does performance-based pay do that in turn improves job performance?

### **Explaining Performance-Based Pay's Effects**

In this section we examine in greater depth the effect of performance-based pay on a number of different employee cognitions. To date there has been little research investigating these factors. We believe that a better understanding of the antecedents of these outcomes will prove helpful in disentangling some of the conflicting findings. In that regard, we will now discuss various effects

performance-based pay may have on employee cognitions and attitudes. We will divide this section into two parts: (a) potentially positive and (b) potentially negative effects of performance-based pay on employee cognitions and perceptions.

#### POSITIVE EFFECTS OF PERFORMANCE-BASED PAY

Based on a review of the research and practice literature, one can group the positive effects of performance-based pay into four categories: increased motivation, improved role clarity, more effective human resource systems, and an increased sense of fairness. Each of these categories is described in greater detail next.

**Increased motivation.** From an expectancy theory perspective, a basic effect we would expect performance-based pay systems to have on employees would be an improvement in perceptions about instrumentalities and expectancies. We believe this is the primary, critical mediator between performance-based pay systems and performance outcomes. The performance-based pay system should increase the likelihood that a “carrot” or valued reward (e.g., a bonus) can be achieved. If the carrot is effective, the possibility of receiving more pay stimulates the employee, and the outcome--it is hoped by designers of performance-based pay systems--is greater effort and better performance. According to expectancy theory, the effects on performance would depend on the valence (i.e., the desirability) of the increased pay, but also on instrumentality (i.e., Can I achieve the goal?) and expectancy (i.e., If I achieve it, will they give me the bonus?) perceptions. While it is unlikely that a performance-based pay program will affect the valence associated with pay, it should affect the instrumentalities and expectancies associated with getting extra pay.

Therefore, a cognition of interest to designers of performance-based pay systems is the future state created in an *employee's mind* when exposed to performance-based pay, as well as feelings or emotions that attach to this potential future state (e.g., how exciting or positive employees imagine the future state to

be). The vision of the potential future state (e.g., higher pay), accompanied by positive emotions upon contemplating it, would promote the incentive effect designers of performance-based pay systems are hoping for.

From a goal-setting perspective, performance-based pay also produces a number of effects. An effective performance-based pay system seems to be able to encourage employees to commit to achieving performance goals. Organizations want employees who are motivated to work hard, but they also want employees to help them achieve specific goals. This is why some organizations provide bonuses for achievement of specific goals. By providing pay contingent on goal attainment, the belief is that employees will be more likely to commit to a goal and devote time and attention to it. As an example, Hughes Aircraft Company used this technique in a big way to deliver on a big contract in 1995. With a \$1 billion contract at stake, Hughes created a \$20 million incentive pool to motivate employees to complete the contract on time (Kausal, 1996). For employees involved in the project, this amounted to almost 75 percent of their annual salary. Hughes was successful, and managers there believe the financial incentives made a significant contribution to the successful effort. Employees overcame many hurdles in the course of completing the project, at times practically living at their workplace to get more done.

Another positive effect of performance-based pay is that it may encourage employees to set spontaneous and higher goals. Some (e.g., Hollensbe, 2000; Wright, 1994) have argued that monetary incentives, particularly when potential payouts are larger, encourage employees to set performance goals when they otherwise wouldn't have. This sounds plausible. We know that goals can have positive effects on performance. Goals “provide us with a clear direction; inform us that we need to try hard; remind us that an end is in sight; and encourage us to think about the process of reaching that end” (Katz, 2000, p. 1). Although the spontaneous goal-setting phenomenon is a relatively new proposition that has not yet received much research, it has been noted. Gellerman (1963), for example, wrote about how some employees turn their work into a kind

of game in striving to reach a certain “score” in terms of units produced or other measures of productivity.

Performance-based pay may not only encourage spontaneous goal setting, but may also encourage employees to set higher goals (Wright, 1994). Moussa (2000), for example, found that piece-rate incentives seemed to encourage subjects in a lab-based study to set difficult goals for themselves and reach higher performance levels. Field studies investigating the interaction of pay systems and self-set employee goals are needed. As expressed by Stajkovic and Luthans (2001), employees “plan courses of action for the future, anticipate the likely consequences of their future actions, and set performance goals for themselves” (p. 583). Performance-based pay can provide a salient reaction to job performance which could encourage employees to be more cognizant, deliberate, and planful.

In sum, performance-based pay systems should generally help employees perceive that their work effort will yield financial rewards in the future (Kahn & Sherer, 1990). As Katz (2000) points out, financial incentives can increase interest in the task and persistence in accomplishing it. The extent to which performance-based pay achieves positive performance outcomes, then, would be mediated by the extent to which the pay system is successful in directing employees’ attention to their work and sustaining their motivation to do it effectively.

**Improved role clarity.** Role clarity, or its opposite (role ambiguity), plays an important role in organizational behavior. We contend that an effective performance-based pay system will help improve role clarity. One way in which this works is that performance-based pay clarifies standards, measures, and job responsibilities. In fact, Streib and Nigro (1993) found that the most frequently cited benefit of performance-based pay among human resource managers was that it helped clarify performance standards and measures. Instituting performance-based pay systems, to stand a chance of being effective, requires a lot of thinking, discussing, and planning. This would result in greater role clarity for employees, thereby reducing stress

and misguided effort, which, in turn, should increase productivity. In some settings, the implementation of performance-based pay is aimed at improving *managerial* performance by requiring managers to do a better job of managing employee performance (Ingraham, 1993). The new pay system will have to be explained to employees, including how it affects their job. These discussions could easily expand to encompass where employees should focus their efforts to achieve top-notch performance. This should result in less role ambiguity, which is known to be a pernicious source of stress and, presumably, diminished productivity.

In addition to spurring management to tighten up operations and clarify roles and responsibilities, performance-based pay is likely to encourage *employees* to seek and share more performance information. The more employees know about the boss’ expectations, the more likely they will be to fulfill them. Consequently, it is in their self-interest to seek out information about what is expected of them. Employees may also be encouraged to provide supervisors with more performance information, such as informing the boss when they encounter performance impediments. This enhanced information flow should have a positive effect on both individual and unit performance.

Finally, performance-based pay systems are useful for role clarity because they provide concrete feedback about performance. Many employees regard the amount of performance feedback they receive as lacking; receiving additional pay is a salient source of performance feedback. Stajkovic and Luthans (2001) refer to this as the informative value of money. They note that the informative content of money is greater when the amount of incentive received is variable based on the level of performance, as in pay-for-performance plans. For example, an excellent performer can be given a raise and told that she received the highest raise in the work group. This is a strong signal that the employee’s contribution was on target. And this may be about the only feedback some employees will receive from their employer.

**More effective human resource management (HRM) systems.** We believe that performance-based pay systems help support

other HRM systems, namely, attraction and retention systems. Proponents of performance-based pay often cite its ability to attract, and especially to retain, the best performers. In terms of attraction, it is often asserted that pay systems can serve as signals of organizational culture that influence who chooses to join an organization. In terms of retention, the belief is that the best performers will leave an organization to work elsewhere where they can earn more money. Performance-based pay reduces the likelihood of best performers leaving by rewarding their excellent performance with monetary rewards. This is another area begging for research attention. There has been some research indicating applicants prefer pay to be fixed (rather than contingent), merit-based, and individual (rather than group) based (Heneman & Judge, 2000). It would be interesting to compare the pay preferences of poor vs. top performers.

In terms of turnover, Kahn and Sherer (1990) raised an interesting issue which we designate as the “discharge-avoidance effect.” If employees perceive that they are paid well compared with other opportunities available to them, they will be motivated to protect their employment. Employees who earn what they consider to be good compensation through a pay-for-performance plan would be motivated not only by the carrot (i.e., potential of future pay increases), but also by the stick (i.e., the loss of their favorable pay situation).

Up till now, the emphasis here has been on performance-based pay as a method for attracting and retaining top performers. But such systems are also potentially a source of pressure on poor performers. Small or non-existent bonuses can cause a poor performer’s pay to actually decrease over time as inflation devalues a fixed salary. Managers hope this will encourage the poor performer to either improve or leave. A common counter-argument is that these unmotivated, poor performers are not likely to leave, knowing that they are not very employable. This, in fact, is an interesting area needing some research attention. What, exactly, are the reactions of poor performers who do not receive incentive pay? It is likely that a sizable proportion of poor performers do not place the

blame on themselves, but instead blame the supervisor or other forces outside themselves.

**Increased sense of fairness.** Finally, it is important to mention that perceptions of effort-reward consonance would be brought about by strong linkages between effort expended and pay received (Miceli, Jung, Near, & Greenberger, 1991). Miceli et al. found that managers’ perceptions of effort-reward consonance were associated with their reactions to their pay system. Effort-reward consonance and its reverse, effort-reward imbalance, are likely to affect much more than just reactions to pay systems, however. Stress researchers, for example, are beginning to understand the negative effects effort-reward imbalance can have on employees. Effort-reward consonance, on the other hand, should encourage employees to continue their efforts and contribute to job satisfaction.

Effort-reward consonance is one facet of equity perceptions. And if we can expect any effect from an effectively functioning performance-based pay system, it would be an increase in equity perceptions. A general expectation is that those who work harder and do more should be paid more, and performance-based pay, at least in theory, fits with this expectation. Any well executed performance-based pay system should be able to deliver this outcome, notwithstanding employees’ self-favoring psychological biases. And there is some evidence that performance-based pay does so. Schay (1988) found evidence that use of performance-based pay increases perceptions of equity among employees. This finding becomes more important when one considers that Schay also found that the strongest predictor of pay satisfaction in her sample of employees was perceived equity. What is not so clear is how low-reward employees perceive things. Do they tend to also have perceptions of equity when they receive smaller performance rewards? The answer undoubtedly would be influenced by a number of person-specific and situation-specific moderators, but it would be nonetheless interesting to know whether, in general, lower-performing employees perceive performance-based pay systems as equitable.

While performance-based pay systems have the potential for creating many positive perceptions, these systems may sometimes have negative effects, which in turn may *reduce* performance. These are discussed in greater detail next.

#### NEGATIVE EFFECTS OF PERFORMANCE-BASED PAY

We can divide these negative effects into three categories: decreased positive motivations, increased negative motivations, and a sense of resentment. We examine these three categories next.

##### **Decreased positive motivation.**

Performance-based pay may reduce intrinsic motivation and motivation for certain important tasks and responsibilities, as well as lower cooperation among employees and departments. In terms of intrinsic motivation, some researchers, chiefly Deci, have argued that extrinsic rewards such as performance-based pay can reduce intrinsic motivation to perform the task. This is because the reward shifts the employee's focus from the task, and one's interest in it, toward the reward that is now contingent on performing the task (Schay, 1988). As Risher (1999) puts it, "it takes the focus away from the motivation derived from the work itself" (p. 298). Although a number of studies have not shown support for Deci's hypothesis (see Bartol & Locke, 2000), his contention should not be dismissed lightly. According to Kohn (1993), who makes a convincing case against reliance on extrinsic reinforcers, "incentive plans do not respond to the extrinsic orientation exhibited by some workers so much as they *create* this focus on financial factors" (p. 141).

Performance-based pay may also reduce the degree of cooperation. Individual-based pay-for-performance plans have been criticized for causing employees to focus on their own performance (and pay) so much that they are less inclined to cooperate with co-workers. The belief is that employees will be so focused on their own outcomes that they will find co-workers' needs and demands as distractions to be avoided. It is possible the same effect could

result from team-based pay-for-performance plans whereby inter-team cooperation could be reduced. On the other hand, organization-wide incentive programs, such as profit-sharing, would reduce these problems. Nevertheless, the possibility that cooperation is reduced could negatively affect performance in an organization.

A performance-based pay plan may also lead to a drop in motivation to perform other desired, but not financially rewarded, work behaviors. It does so by directing employee behavior toward some desirable behaviors at the expense of others. Wright (1994) called this the "goal only" strategy whereby employees focus too much on attainment of selected goals at the expense of other important work behaviors. For example, employees receiving commission for shoe sales may shortchange the organization's overall effectiveness because other important behaviors are not explicitly rewarded. Employees may avoid training new salespeople or may not take the time to organize the shoes and carefully return them to their proper location. This single-mindedness on the part of employees may not be a problem in some cases, but in others, particularly beyond short-lived incentive plans, it probably will be. Organizations need employees to engage in a variety of positive work behaviors, far too many to encompass in performance-based pay plans.

Writers often decry performance-based pay's effects on extra-role behavior or organizational citizenship behavior (Lee, 2001). There is some evidence that performance-based pay can discourage positive extra-role behaviors. Deckop and Mangel (1999), for example, found that performance-based pay was negatively related to organizational citizenship behavior for some employees. Therefore, pay-system designers will need to take care to create pay systems that do not cause positive behaviors to disappear.

Finally, Wright (1994) has noted something he calls the "easy goal" effect. This is the problem of employees being less than forthcoming about their true performance capabilities so they can set goals that will be easier to attain. This has been written about much in terms of blue-collar employees who

hide their true capabilities in piece-work systems, but it happens in white-collar environments as well. Employees in management-by-objectives systems may understandably want a bit of slack in terms of an objective so they won't be pressed to the wall in attempting to reach it. Basing pay on reaching a given objective builds in a penalty for not reaching it; therefore, employees have an incentive to hide their true capabilities. This may lead to suboptimization of performance as borne out by stories of newly trained temporary workers who far exceed normal production output in piecework settings.

**Increased negative motivation.** By negative motivation we mean that performance-based pay may encourage employees to engage in certain behaviors that are counterproductive to the organizational well-being. For example, some employees will engage in unethical behavior to attain a goal and receive resulting financial rewards. Wright (1994) calls this the "end justifies the means" effect. The cognitive effect going on in this case is not completely clear, but it seems to be that employees find the valence of the rewards outweigh the risk-adjusted consequences of getting caught. We will call the cognitive precursor of such unethical behavior the "temptation effect." A widely known case of this was the Sears auto repair fiasco, where customers were billed for unneeded repairs so employees could collect incentive pay. In other cases, it is simply that payouts are determined in a way that encourages behavior that seems unethical. For example, some insurance companies have been criticized for awarding bonuses tied to reducing claims payouts (Frey, 2000). Organizations obviously need to do a better job of predicting undesirable behaviors that a proposed pay plan may encourage.

Performance-based pay may also increase destructive competition among employees. As individual employees or teams of employees pursue their objectives and seek to maximize their financial rewards, overall organizational performance may be damaged by employees competing in a way that harms the organization. The classic example might be salespeople fighting for customers in such a way that the customers leave the store and nobody

gets the sale. These variables need to be studied more carefully.

Finally, performance-based pay may also create and encourage a short-term orientation. Pearce (1987) asserts that individual-based rewards distract employees' attention away from the fact that they are in a situation of "pooled interdependence" (p. 522). Employees lose sight of the fact that maintaining their employment and steady income stream depends on the continuing existence of their employer. This comes about because specifying pay-for-performance parameters creates a "pseudocontract" between buyer (employers) and seller (employees) that, "can, at best, cover only a portion of the desired actions and becomes a forced and artificial representation of the kind of performance that would be most effective for the organization" (Pearce, p. 523). Employees develop a short-term, transactional mentality that is more like a contractor's than an employee's. And this contractor mentality, according to Pearce, reduces organizational commitment and undermines the employee-employer relationship.

**Greater sense of resentment.** We think most proponents of performance-based pay would agree that its intention is to shape employee behavior through the use of financial rewards. However, to some employees, this shaping may feel like manipulation, particularly in more salient plans where rewards and their preconditions are harder to ignore. Some (e.g., Kohn, 1993) even go so far as to say that rewards feel punitive because they are manipulative. Deci (2001, personal communication) is of the opinion that using performance-based rewards can be like "playing with fire, because they very easily spill over into the realm of control, both in the way they are used and the way they are experienced." Some employees have perceived performance-based pay programs as attempts to control their behavior and encourage conformity (Siegall & Worth, 2001). This perception would likely result in feelings of resentment and negative attitudes. A challenge, then, seems to be to encourage and reward positive work behavior without creating an overly controlling work context.

Anger can also result from unmet expectations, such as when pay increases are smaller than expected (Archer, 2001). And sometimes anger erupts over others' financial rewards, as in the case of some of the obscenely high executive bonuses regularly reported in the media. General Motors provided a memorable example when, while closing down plants because of reduced demand, it paid huge performance bonuses to its top executives, much to the chagrin of the United Auto Workers and the displaced workers.

In sum, performance-based pay systems may affect performance in a wide variety of ways. To our knowledge, there has been almost no research studying several of these effects simultaneously. More studies are needed to better understand the effects of such systems.

### CONCLUSION

To summarize, research on performance-based pay suggests that it has a positive effect on performance. Nevertheless, as Gellerman pointed out, "the impact of money is very complex and easy to misunderstand" (1963, p. 67). Our hope is that the effects we have described above will help sort out this messy subject. Knowing more about these effects should help practitioners design programs that minimize negatives and maximize positives. So should knowing more about actual, as opposed to intended, effects of performance-based pay. Managers and researchers need to spend more time in the field asking employees how a given pay system affects them. Only then can we see how well our theories and practices are actually borne out in the workplace.

### REFERENCES

- Archer, J. (2001). Smaller-than-expected bonuses anger some L.A. teachers. *Education Week*.  
<http://www.edweek.org/ew/newstory.cfm?slug=10bonus.h21>
- Barnard, C. I. (1938). *The functions of the executive*. Cambridge, MA: Harvard University.
- Bartol, K. M., & Locke, E. A. (2000). Incentives and motivation. In S. L. Rynes and B. Gerhart (Eds.), *Compensation in organizations: Current research and practice*, (pp. 104-147). San Francisco: Jossey-Bass.
- Collins, J. (1999). Turning goals into results: The power of catalytic mechanisms. *Harvard Business Review*, 77(4), 70-83.
- Deckop, J. R., & Mangel, R. (1999). Getting more than you pay for: Organizational citizenship behavior and pay-for-performance plans. *Academy of Management Journal*, 42, 420-429.
- Flannery, T. P., Hofrichter, D. A., & Platten, P. E. (1996). *People, performance, and pay: Dynamic compensation for changing organizations*. New York: The Free Press.
- Frey, J. (2000). Insurance adjusters rewarded for shrinking claims checks. *Insure.com*.  
<http://www.insure.com/gen/adjusterperformance.html>
- Fried, Y. (1991). Meta-analytic comparison of the job diagnostic survey and job characteristics inventory as correlates of work satisfaction and performance. *Journal of Applied Psychology*, 76, 690-697.
- Gellerman, S. W. (1963). *Motivation and productivity*. United States: American Management Association.
- Harris, M. (2000). *Human resource management: A practical approach* (2<sup>nd</sup> ed.). Forth Worth, TX: Dryden.
- Heneman, H. G., III, & Judge, T. A. (2000). Compensation attitudes. In *Compensation in organizations: Current research and practice*, pp. 61-103, S. L. Rynes & B. Gerhart (Ed.). San Francisco: Jossey-Bass.
- Hollensbe, E. C. (2000). Group pay for performance plans: The role of spontaneous goal setting. *Academy of Management Review*, 25, 864-873.
- Ingraham, P. W. (1993). Pay for performance in the states. *American Review of Public Administration*, 23, 189-200.
- Jenkins, G. D. Jr., Mitra, A., Gupta, N., & Shaw, J. D. (1998). Are financial incentives related to Performance? A meta-analytic review of empirical research. *Journal of Applied Psychology*, 83, 777-787.
- Kahn, L. M., & Sherer, P. D. (1990). Contingent pay and managerial performance. *Industrial and Labor Relations Review*, 43, 107-S-120-S.
- Katz, N. R. (2000). Incentives and performance management in the public sector. *Public sector performance management executive session: An initiative of the vision of governance in the twenty-first century project*.  
[http://www.ksg.harvard.edu/visions/performance\\_management/katz\\_incentives.htm](http://www.ksg.harvard.edu/visions/performance_management/katz_incentives.htm)
- Kausal, B. A. (1996, March-April). Peace Shield—a study in motivation. *PM*.
- Kohn, A. (1993). *Punished by rewards: The trouble with gold stars, incentive plans, A's, praise, and other bribes*. Boston: Houghton Mifflin.
- Lee, S. (2001). Does performance-base pay motivate? *PA Times*, 24(4), 3.
- Miceli, M. P., Jung, I., Near, J. P., & Greenberger, D. B. (1991). Predictors and outcomes of reactions to pay-for-performance plans. *Journal of Applied Psychology*, 76, 508-521.
- Mitchell, T.R. (1999). The meaning of money: An individual-difference perspective. *Academy of Management Review*, 24, 568-579.
- Moussa, F. M. (2000). Determinants, process, and consequences of personal goals and performance. *Journal of Management*, 26, 1259-1286.

- Opsahl, R. L. (1967). Managerial compensation: Needed research. *Organizational Behavior and Human Performance*, 2, 208-216.
- Pearce, J. L. (1987). Why merit pay doesn't work: Implications from organizational theory. In R. M. Steers, L. W. Porter, & G. A. Bigley (Eds.), *Motivation and leadership at work* (6<sup>th</sup> ed.), (pp. 519-526). New York: McGraw-Hill.
- Risher, H. (1999). Using pay as a tool to achieve organizational goals. In H. Risher (Ed.), *Aligning pay and results: Compensation strategies that work from the boardroom to the shop floor*. (pp. 295-311). New York: AMACOM.
- Schay, B. W. (1988). Effects of performance-contingent pay on employee attitudes. *Public Personnel Management*, 17, 237-250.
- Siegall, M., & Worth, C. (2001). The impacts of trust and control on faculty reactions to merit pay. *Personnel Review*, 30, 646-656.
- Stajkovic, A. D., & Luthans, F. (2001). Differential effects of incentive motivators on work performance. *Academy of Management Journal*, 44, 580-591.
- Streib, G., & Nigro, L. G. (1993). Pay for performance in local governments: Programmatic differences and perceived utility. *Public Productivity & Management Review*, 17, 145-159.
- Wang, L., Kick, E., Fraser, J., & Burns, T. J. (1999). Status attainment in America: The roles of locus of control and self-esteem in educational and occupational outcomes. *Sociological Spectrum*, 19, 281-299.
- Wiscombe, J. (2001). Can pay for performance really work? *Workforce*, 80(8), 28-33.
- Wright, P. M. (1994). Goal setting and monetary incentives: Motivational tools that can work *too* well. *Compensation and Benefits Review*, May-June, 41-49.

**Authors' Note:** Correspondence may be addressed to: Brad Gilbreath, Indiana University-Purdue University Fort Wayne, Organizational Leadership & Supervision, 288 Neff Hall, Fort Wayne, IN 46805. Email address: [gilbreaj@ipfw.edu](mailto:gilbreaj@ipfw.edu) or Michael M. Harris, College of Business Administration, University of Missouri-St. Louis, St. Louis MO 63121. Email address: [nharris@umsl.edu](mailto:nharris@umsl.edu)



## THE ESTIMATION OF INTEROBSERVER AGREEMENT IN BEHAVIORAL ASSESSMENT

April A. Bryington, Darcy J. Palmer, and Marley W. Watkins  
The Pennsylvania State University

Direct observation of behavior has traditionally been a core component of behavioral assessment. However, systematic observational data is not intrinsically reliable and valid. It is well known that observer accuracy and consistency can be influenced by a variety of factors. Therefore, interobserver agreement is frequently used to quantify the psychometric quality of behavioral observations. Two of the commonly used interobserver agreement indices, percentage of agreement and kappa, are reviewed. Although percentage agreement is popular due to its computational simplicity, kappa has been found to be a superior measure because it corrects for chance agreement among observers and allows for multiple observers and categories. A description of kappa and computational methods are presented.

Direct observation of behavior has traditionally been a core component of behavioral assessment (Ciminero, 1986; Tryon, 1998). Originally, it was thought unnecessary to establish the reliability and validity of direct observations of behavior since by definition direct observation is free of bias and valid. However, various aspects of methodology can confound the data and therefore lead to invalid results (Hops, Davis, & Longoria, 1995).

Kazdin (1977) reviewed research that demonstrated that observer accuracy and reliability can be influenced by variables such as knowledge that accuracy is being checked, drift from original definitions of the observed behavior, the complexity of the coding system being used, and observer expectancies combined with feedback. In addition, Wasik and Loven (1980) reported that characteristics of the recording procedures, characteristics of the observer, and characteristics unique to the observation setting are sources of inaccuracy that can jeopardize the reliability and validity of observational data. Consequently, Cone (1998) suggested that the quality of any observations of behavior must be determined regardless of the procedures used to quantify them.

### INTEROBSERVER AGREEMENT

Researchers have identified procedures that can be used to measure the psychometric properties of data obtained from direct observation (Primavera, Allison, & Alfonso, 1997). The most common of these procedures is interobserver agreement (Skinner, Dittmer, & Howell, 2000). There are diverse opinions of what interobserver agreement actually measures.

Hops et al. (1995) defined interobserver agreement as a measure of consistency and, therefore, as representing a form of reliability. In contrast, Alessi (1988) described interobserver agreement as an estimate of objectivity that indicates the degree to which the data reflect the behavior being observed rather than the behavior of the observer. Alessi's definition implies that interobserver agreement is tapping into aspects of validity. Suen (1988, 1990) indicated that interobserver agreement could serve as a measure of both reliability and validity depending upon the degree to which two or more observers agree on occurrences or nonoccurrences, whether a criterion-referenced or norm-referenced orientation is used, and the ratio of random to systematic error. Although there are divergent views about what agreement actually measures, it is generally accepted that it is fundamental to sound behavioral measurement for both researchers and practitioners (Bloom, Fischer, & Orme, 1999; Hayes, Barlow, & Nelson-Gray, 1999; Hoge, 1985; Hops et al., 1995; Kazdin, 2001; Kratochwill, Sheridan, Carlson, & Lasecki, 1999; Maag, 1999; McDermott, 1988; Salvia & Ysseldyke, 2001; Suen, 1988).

### Assessing Interobserver Agreement

Many different methods of calculating interobserver agreement have been proposed (Berk, 1979; Hartmann, 1977; House, House, & Campbell, 1981; Shrout, Spitzer, & Fleiss, 1987). The two most commonly cited methods are percent of agreement and kappa.

### Overall Percent of Agreement

The most frequently used method for determining interobserver agreement is overall percent of agreement (Berk, 1979; Hartmann, 1977; McDermott, 1988). Percent of agreement is calculated by the benefits of using overall percent of agreement include its ease of calculation and interpretation (Hartmann). The disadvantages of percent of agreement, however, have caused many researchers to caution against its use (Berk; Birkimer & Brown, 1979; Hartmann; Hops et al., 1995; McDermott; Shrout et al., 1987; Suen & Lee, 1985; Towstapiat, 1984).

The most significant problem with percent of agreement is its failure to take into account agreement that may be due to chance (House et al., 1981). As McDermott (1988) pointed out, when using percent of agreement “there exists no means of determining whether obtained agreement is effectively beyond what might be produced by completely naive observers or by the toss of dice” (p. 229). Not only does percent of agreement fail to control for chance, it is also influenced by the frequency of behaviors being observed. A researcher may obtain a level of percentage agreement that he or she feels is adequate, when in reality, it may be inflated due to chance or the high frequency of the behavior being observed (Towstapiat, 1984). Figure 1 illustrates this potential inflation with data from House, Farber, and Nier (1983).

Suen and Lee (1985) provided empirical evidence that disregarding chance can lead to inflated levels of agreement. They randomly selected 12 studies that reported percentage agreement. From these studies, they chose a simple random sample of 50 observation points and found that between one-fourth and three-fourths of the observations would have been determined to be unreliable against a lenient chance-corrected criterion. Between one-half and three-fourths of the observations would have been judged unreliable against a more stringent chance-corrected criterion. Suen and Lee concluded that percent of agreement has seriously undermined the reliability of past observations and that “its continued use can no longer be justified” (p. 232).

### Occurrence and Nonoccurrence Percent of Agreement

The failure of overall percent of agreement to take chance into account can be partially corrected by using percent of agreement only on the occurrence or nonoccurrence of the target behavior rather than the overall level of agreement. If the occurrence of the target behavior is the focus of interest then percent of agreement on occurrence of the target behavior may be appropriate. Conversely, if agreement on nonoccurrence is most important then percent of agreement on nonoccurrence of the target behavior can be used. These indices indicate the percentage of time in which two or more observers agree that a target behavior either occurred or did not occur.

The benefits of percent agreement on occurrence or nonoccurrence are simplicity of calculation and partial resistance to the distorting effects of chance. However, they do not completely control for chance (Hopkins & Herman, 1977) and they can potentially produce incongruent indices of agreement. Like overall percent of agreement, percent agreement on occurrence or nonoccurrence is only applicable when two observers are monitoring a dichotomous target behavior (Primavera et al., 1997).

### Kappa Coefficient of Agreement

Kappa ( $k$ ; Cohen, 1960) has become the preferred index for measuring interobserver agreement (Hops et al., 1995). For example, Primavera et al. (1997) highly recommended kappa “when data are dichotomous or nominal” (p. 64). While Langenbucher, Labouvie, and Morgenstern (1996) suggested that kappa “should be the default measure” (p. 1287) when assessing diagnostic agreement in psychiatry. Kappa has also been favored for determining observer agreement in medicine (Everitt, 1994).

### Strengths of kappa

One of kappa’s strengths is its ability to correct for chance agreement across two or more nominal categories. Another is its known sampling distribution that allows for the construction of confidence intervals and tests of statistical significance (Cohen, 1960). An

original limitation of kappa was that it could only be used with two observers and the same two observers had to rate every observation. This was corrected by Fleiss (1971) who extended kappa to be used in situations in which there are a constant number of raters, but the raters do not necessarily have to be the same across observations. Fleiss's  $k_m$  (the subscript  $m$  signifying  $k$  for multiple observers) automatically reduces to  $k$  when there are only two observers for all observations.

Another beneficial characteristic of kappa is that it allows for generalizability across different experimental conditions. Foster and Cone (1986) pointed out that chance agreement changes as the base rate or prevalence of behavior changes. Because percent of agreement does not correct for chance, it is differentially inflated in situations with different rates of behavior, hindering comparison across conditions. Kappa, however, allows for standardized comparisons by statistically removing chance.

### Limitations of kappa

Although kappa's benefits have caused many to suggest that it is the most desirable index to use when calculating interobserver agreement, it also has several limitations that should be considered. One constraint of kappa is that it can only be used with nominal scale data. Because most interobserver comparisons involve nominal categorization, this is generally not a problem. A second possible limitation is that kappa is impossible to calculate when both observers report that the behavior occurred 100% of the time or not at all. When this occurs, chance agreement will equal 100% and the denominator of the kappa equation will resolve to zero (Foster & Cone, 1986). However, this is more of a theoretical problem than a practical one. If observers agree 100% of the time, it can be seen as perfect agreement.

Another possible limitation of the kappa coefficient is that it tends to decrease when there are low base rates of the observed behavior (Shrout et al., 1987). To alleviate this problem, Nelson and Cicchetti (1995) suggested that researchers ensure that there are at least ten occurrences of the behavior in the sample being

observed. This will minimize the effect of interobserver disagreement in cases of low frequency behaviors. Similarly, the magnitude of kappa can be influenced by the relative balance of agreements and disagreements. However, Cicchetti and Feinstein (1990) pointed out that this tendency serves a legitimate scientific purpose.

### Interpretation of kappa

Kappa indicates the proportion of agreement above and beyond what would be expected by chance (Cohen, 1960) and takes the form of a simple correlation coefficient that is relatively easy to interpret. Possible values range from +1.00, which indicates perfect agreement, through 0.00, which reflects chance agreement, down to a theoretical -1.00, which signifies perfect disagreement. Values less than zero are usually of no practical interest because they represent agreement that is less than would be expected by chance (Cohen). Because kappa adjusts for chance agreement, less stringent guidelines are generally applied than those used in simple percent of agreement. Cicchetti (1994) provided a summary of interpretive guidelines for kappa. Specifically, values below 0.40 indicate poor clinical significance; values between 0.40 and 0.59 indicate fair clinical significance; values between 0.60 and 0.74 indicate good clinical significance; and values between 0.75 and 1.00 indicate excellent clinical significance. Because kappa accounts for chance, a coefficient of +1.00 can be interpreted correctly as indicating perfect agreement between observers. In this case, the observers would have accounted for 100% of the agreement that was not explained by chance. If a coefficient of zero is obtained, it indicates that the observers' ratings are no more precise than what could be attained by random assignment. A kappa coefficient of 0.80 indicates that the observers have accounted for 80% of the agreement over and above what would be expected by chance.

### Calculation of kappa

Conceptually, kappa is defined as: The greatest deterrent to the use of kappa may be its perceived difficulty of computation when compared to simple percent agreement (Hops et

al., 1995). Therefore, this paper presents two methods to simplify the calculation of kappa. The first method is appropriate for the case of two observers and is easily computed by hand. An algorithm and sample calculation are provided in Figure 1. A REALbasic computer program, entitled *Chi-Square Analysis* (Watkins, 2002), is also available for the case of two observers. Both Macintosh and Windows versions can be downloaded without charge from <http://espse.ed.psu.edu/spsy/Watkins/SPSY-Watkins.ssi>.

The second method is more complex and therefore must be automated with a computer. It is based upon the Fleiss (1971) extension of kappa to the case of multiple observers, where the observers do not have to remain constant throughout the study. This computer program, entitled *MacKappa* (Watkins, 1998), calculates partial kappa coefficients to allow the investigator to verify agreement on a category-by-category basis as well as by the overall weighted average across categories. It also provides sampling distribution data to allow the researcher to ascertain the statistical significance of general and partial kappa coefficients. *MacKappa* is a FutureBASIC program that operates on Macintosh computers under Mac OS 9. Data is input via a tab delimited text file. *MacKappa* will conduct analyses with 2-999 observers, 2-999 cases, and 2-25 categories. *MacKappa* can be downloaded without charge from <http://espse.ed.psu.edu/spsy/Watkins/SPSY-Watkins.ssi>.

#### SUMMARY

The calculation of interobserver agreement is essential for establishing the psychometric properties of observational data. Although percentage agreement is the most commonly used agreement index, its limitations have led researchers to recommend kappa as a more desirable index of interobserver agreement. Difficult computation may have deterred its common use in the past; however, this is no longer a salient problem with the computational guide and computer programs presented in the current paper.

#### REFERENCES

- Alessi, G. (1988). Direct observation methods for emotional/behavioral problems. In E. S. Shapiro & T. R. Kratochwill (Eds.), *Behavioral assessment in schools: Conceptual foundations and practical applications* (pp. 14-75). New York: Guilford Press.
- Berk, R. A. (1979). Generalizability of behavioral observations: A clarification of interobserver agreement and interobserver reliability. *American Journal of Mental Deficiency, 83*, 460-472.
- Birkimer, J. C., & Brown, J. H. (1979). Back to basics: Percentage agreement measures are adequate, but there are easier ways. *Journal of Applied Behavior Analysis, 12*, 535-543.
- Bloom, M., Fischer, J., & Orme, J. G. (1999). *Evaluating practice: Guidelines for the accountable professional* (3<sup>rd</sup> ed.). Boston: Allyn and Bacon.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*, 284-290.
- Cicchetti, D. V., & Feinstein, A. R. (1990). High agreement but low kappa: II. Resolving the paradoxes. *Journal of Clinical Epidemiology, 43*, 551-558.
- Ciminero, A. R. (1986). Behavioral assessment: An overview. In A. R. Ciminero, K. S. Calhoun, & H. E. Adams (Eds.), *Handbook of behavioral assessment* (2<sup>nd</sup> ed., pp. 3-11). New York: John Wiley & Sons.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46.
- Cone, J. D. (1998). Psychometric considerations: Concepts, contents, and methods. In A. S. Bellack & M. Hersen (Eds.), *Behavioral assessment: A practical handbook* (4<sup>th</sup> ed., pp. 22-46). Boston: Allyn & Bacon.
- Everitt, B. S. (1994). *Statistical methods in medical investigations* (2<sup>nd</sup> ed.). London: Edward Arnold.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin, 76*, 378-382.
- Foster, S. L., & Cone, J. D. (1986). Design and use of direct observation procedures. In A. R. Ciminero, K. S. Calhoun, & H. E. Adams (Eds.), *Handbook of behavioral assessment* (2<sup>nd</sup> ed., pp. 253-324). New York: John Wiley & Sons.
- Hartmann, D. P. (1977). Considerations in the choice of interobserver reliability estimates. *Journal of Applied Behavior Analysis, 10*, 103-116.
- Hayes, S. C., Barlow, D. H., & Nelson-Gray, R. O. (1999). *The scientist-practitioner: Research and accountability in the age of managed care*. Boston: Allyn and Bacon.
- Hoge, R. D. (1985). The validity of direct observational measures of pupil classroom behavior. *Review of Educational Research, 55*, 469-483.
- Hopkins, B. L., & Hermann, J. A. (1977). Evaluating interobserver reliability of interval data. *Journal of Applied Behavior Analysis, 10*, 121-126.
- Hops, H., Davis, B., & Longoria, N. (1995). Methodological issues in direct observation: Illustrations with the living in familial environments (LIFE) coding system. *Journal of Clinical Child Psychology, 24*, 193-203.

- House, A. E., Farber, J. W., & Nier, L. L. (1983). Differences in computational accuracy and speed of calculation between three measures of interobserver agreement. *Child Study Journal, 13*, 195-201.
- House, A. E., House, B. J., & Campbell, M. B. (1981). Measures of interobserver agreement: Calculation formulas and distribution effects. *Journal of Behavioral Assessment, 3*, 37-57.
- Kazdin, A. E. (1977). Artifact, bias, and complexity of assessment: The ABC's of reliability. *Journal of Applied Behavior Analysis, 10*, 141-150.
- Kazdin, A. E. (2001). *Behavior modification in applied settings* (6<sup>th</sup> ed.). Belmont, CA: Wadsworth/Thomson Learning.
- Kratochwill, T. R., Sheridan, S. M., Carlson, J., & Lasecki, K. L. (1999). Advances in behavioral assessment. In C. R. Reynolds & T. R. Gutkin (Eds.), *The handbook of school psychology* (3<sup>rd</sup> ed.) (pp. 350-382). New York: Wiley.
- Langenbucher, J., Labouvie, E., & Morgenstem, J. (1996). Measuring diagnostic agreement. *Journal of Consulting and Clinical Psychology, 64*, 1285-1289.
- Maag, J. W. (1999). *Behavior management: From theoretical implications to practical applications*. San Diego, CA: Singular Publishing.
- McDermott, P. A. (1988). Agreement among diagnosticians or observers: Its importance and determination. *Professional School Psychology, 3*, 225-240.
- Nelson, L. D., & Cicchetti, D. V. (1995). Assessment of emotional functioning in brain-impaired individuals. *Psychological Assessment, 7*, 404-413.
- Primavera, L. H., Allison, D. B., & Alfonso, V. C. (1997). Measurement of dependent variables. In R. D. Franklin, D. B. Allison, & B. S. Gorman (Eds.), *Design and analysis of single-case research* (pp. 41-91). Mahwah, NJ: Erlbaum.
- Salvia, J., & Ysseldyke, J. E. (2001). *Assessment* (8<sup>th</sup> ed.). Boston: Houghton Mifflin.
- Shrout, P. E., Spitzer, R. L., & Fleiss, J. L. (1987). Quantification of agreement in psychiatric diagnosis revisited. *Archives of General Psychiatry, 44*, 172-177.
- Skinner, C. H., Dittmer, K. I., & Howell, L. A. (2000). Direct observation in school settings: Theoretical issues. In E. S. Shapiro & T. R. Kratochwill (Eds.), *Behavioral assessment in schools: Theory, research, and clinical foundations* (2<sup>nd</sup> ed., pp. 19-45). New York: Guilford Press.
- Suen, H. K. (1988). Agreement, reliability, accuracy, and validity: Toward a clarification. *Behavioral Assessment, 10*, 343-366.
- Suen, H. K. (1990). *Principles of test theories*. Hillsdale, NJ: Erlbaum.
- Suen, H. K., & Lee, P. S. C. (1985). Effects of the use of percentage agreement on behavioral observation reliabilities: A reassessment. *Journal of Psychopathology and Behavioral Assessment, 7*, 221-234.
- Towstoptiat, O. (1984). A review of reliability procedures for measuring observer agreement. *Contemporary Educational Psychology, 9*, 333-352.
- Tryon, W. W. (1998). Behavioral observation. In A. S. Bellack & M. Hersen (Eds.), *Behavioral assessment: A practical handbook* (4<sup>th</sup> ed., pp. 79-103). Boston: Allyn & Bacon.
- Wasik, B. H., & Loven, M. D. (1980). Classroom observational data: Sources of inaccuracy and proposed solutions. *Behavioral Assessment, 2*, 211-227.
- Watkins, M. W. (1988). *MacKappa* [Computer software]. Pennsylvania State University: Author.
- Watkins, M. W. (2002). *Chi-Square Analysis* [Computer software]. Pennsylvania State University: Author.

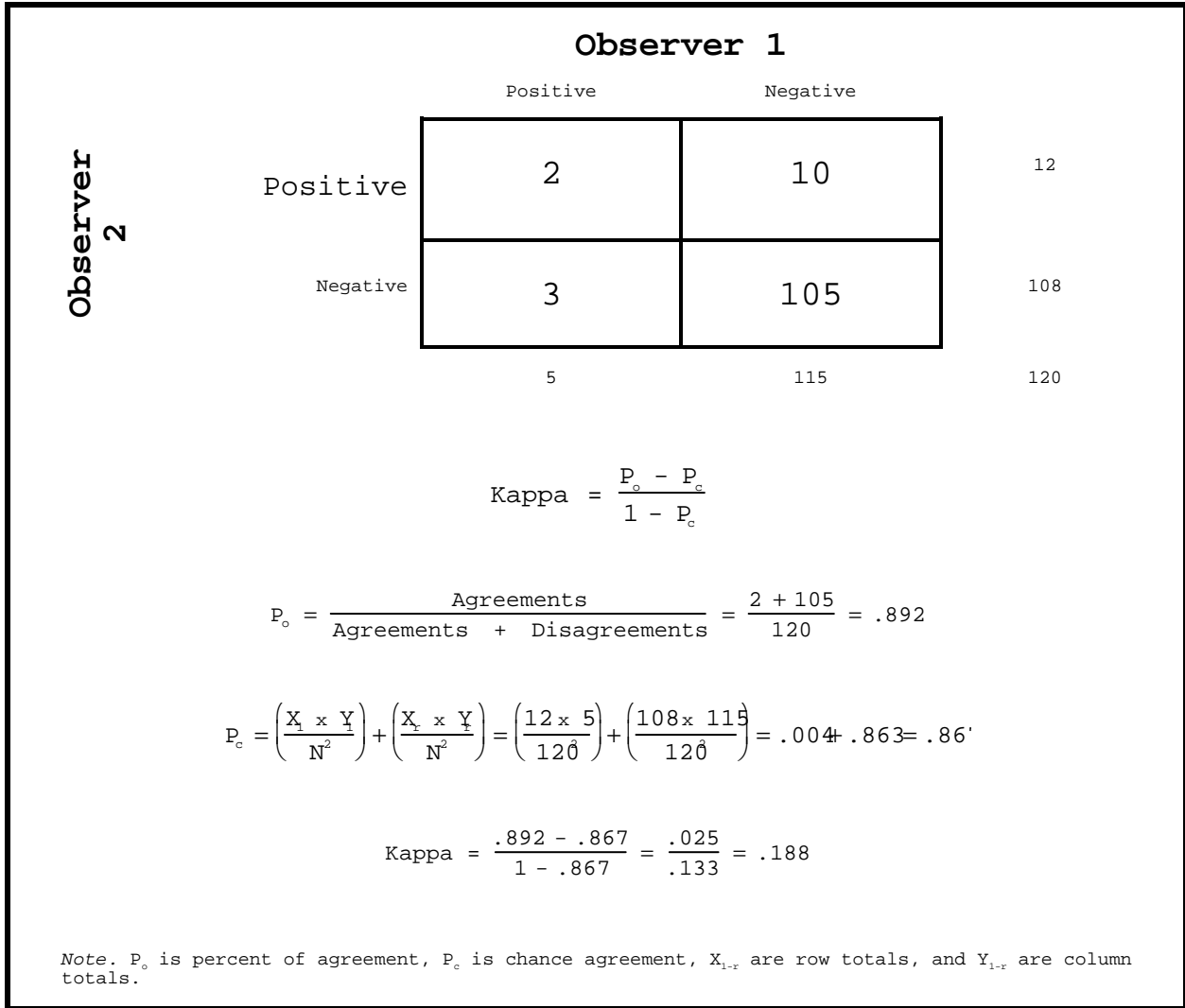


Figure 1. Algorithm and sample calculation for Kappa for two observers who rate 120 cases into two mutually exclusive categories.

## USING PEERS TO FACILITATE LEARNING BY STUDENTS WITH MODERATE DISABILITIES

Belva C. Collins  
University of Kentucky

Based on the research of the author with her colleagues, this article describes the effective use of peers in systematic procedures to facilitate learning by students with moderate disabilities. Five research investigations are described, and guidelines for using on peers are offered. Specifically, the article addresses the use of peers in the direct delivery of systematic procedures in both special and general education classrooms to facilitate acquisition and generalization. In addition, guidelines for using peers in the delivery of systematic procedures are offered.

---

Systematic instruction (Wolery, Ault, & Doyle, 1992) is an effective means of teaching skills to students with moderate disabilities. Systematic instructional trials consist of the delivery of an attentional cue, a discriminative stimulus (e.g., task direction), a delay interval, a prompt, a response interval, and a consequence. Response prompting strategies are a form of systematic instruction that employ prompts to increase the likelihood of correct responses. Examples include the time delay procedure in which the controlling prompt is faded over time as students begin to respond correctly before the prompt and the system of least prompts in which prompts are systematically faded by offering only the amount of assistance necessary from a set prompt hierarchy. To increase the efficiency of a response, the instructor may choose to add nontargeted information (i.e., additional information in the antecedent, the prompt, or the consequence) that may be acquired through repeated exposure over time. Also, the instructor may choose to use additional strategies that have been shown to facilitate generalization (e.g., multiple exemplars of material, instructors, or settings).

One way that instructors can increase the number of instructional trials per day is to involve peers without disabilities. Peers are readily available to participate in instruction within the general education settings, and many special education classrooms for students with disabilities employ peer tutors within that setting. The purpose of this article is to describe a series of research investigations focusing on peer-mediated strategies that were conducted by the author and her colleagues over the course of several years. Specifically, the article will describe six research studies conducted in special and general education settings in which

peers served as tutors in the direct delivery of instructors or as assistants to facilitate generalization and the acquisition of nontargeted information. In addition, the author discusses guidelines for using peers in systematic interactions with students with moderate disabilities

### REVIEW OF SELECTED RESEARCH STUDIES

The following studies focused on the use of peers in the delivery of systematic procedures to facilitate learning. The studies all involved students with functional mental disabilities (FMD), which is the label Kentucky uses to secure services for students with moderate to severe disabilities, and their same-age peers without disabilities. In addition, each study employed a single subject research design to investigate a different aspect of peer involvement. An overview of the investigation is shown in Table 1, and a summary of each investigation is described in the following sections.

#### **Peers as Reliable and Effective Instructors**

Several of the investigations focused on the use of peers to deliver direct systematic instruction. In addition to addressing several research questions, each investigation examined whether peers can reliably implement systematic instructional procedures and whether that instruction is effective in teaching target behaviors.

*Peer- vs. teacher-delivered instruction in the special education classroom.* Miracle, Collins, Schuster, and Grisham-Brown (2001) compared peer- to teacher-delivered instruction in teaching functional sight words to 4 students with FMD. In this investigation, the special

education teacher selected a constant time delay procedure because it has been an effective procedure for peer tutors to use to teach sight words to students with disabilities (Koury & Browder, 1986) and because it was a procedure she typically used in her classroom.

To train the 5 peer tutors in the constant time delay procedure, the teacher conducted 30-minute training sessions in the morning before school for 3 days. The training sessions consisted of (a) an overview of disabilities, the time delay procedure, and the investigation; (b) modeling of the time delay and data collection procedures; and (c) practice using the time delay procedure and collecting data until the peer tutors met a procedural reliability criterion of 86% correct responses. Once peer training was complete, Miracle et al. (2001) used an alternating treatments design to compare the effectiveness and efficiency of teacher- versus peer-delivered instruction to teach functional sight words to the students with FMD. Through a screening process, the teacher selected 15 words and divided them into three counterbalanced sets of five words each: (a) one set to be taught by the teacher, (b) one set to be taught by the peers, and (c) one set to be used as a control set.

Following baseline condition, teacher- and peer-delivered instruction began. Each day, the students with disabilities participated in one instructional session conducted by the teacher, one instructional session conducted by a peer tutor, and one probe session conducted by the teacher on the control set. Each session consisted of two trials per target word, with the order of the targeted words rotated daily. A trained observer collected both interobserver (i.e., dependent variable) and procedural (i.e., independent variable) reliability data once per condition per student on each set of words.

The results showed the constant time delay procedure was effective in teaching 7 out of 8 sets of target words to 3 of the 4 participants. Procedural reliability agreement averaged 100% for the teacher and 98.3% for the peer tutors. Only one peer tutor made an error and was retrained when this occurred. Interobserver reliability agreement was 100% for both teacher- and peer-delivered instruction.

Although differences were minimal, the students with disabilities met criterion during teacher-delivered instruction before they did in peer-delivered instruction. Students maintained the words they were taught with 80-100% accuracy over time.

*Peer delivery of targeted and nontargeted information in the special education classroom.* Collins, Branson, and Hall (1995) selected 26 students without disabilities from an Advanced English class to teach key words from cooking product labels to 4 students with FMD in a special education classroom setting. Using this group of peer tutors had several advantages. First, the peer tutoring project gave the students in advanced classes the opportunity to interact with students with disabilities who typically were not included in their coursework. Second, the peer tutoring project resulted in written portfolio entries for the students without disabilities and alternate portfolio entries for the students with FMD, as required under the Kentucky Education Reform Act of 1990 (Kleinert, Kearns, & Kennedy, 1997).

In a training session in the English class, the special education teacher presented an overview of the disabilities and the constant time delay procedure. The university investigator then taught the peer tutors to conduct probe and instructional sessions and to collect and graph daily data. She did this through (a) providing written instructions, (b) giving an oral presentation, (c) modeling the procedure, (d) conducting role-play with feedback, and (e) answering students' questions. She then gave a written quiz over the procedures and gave the peer tutors a condensed set of written instructions to use during their sessions. The university investigator worked with the English teacher in assigning specific peer tutors to work with specific students with FMD. Prior to intervention, the English teacher had peer tutors anonymously write their reactions to students in the FMD classroom.

Once the university investigator selected the target words and nontargeted information, the peer-tutoring project began. Each day, two peer tutors worked with one of the students with FMD in the special education classroom. One peer tutor conducted the instruction while the



other collected data. During instructional sessions, a peer tutor showed a flash card containing a target word to a student and asked the student to locate it on the labels on one of two brands of a cooking product. Once the student pointed to the word, the peer tutor used the time delay procedure to prompt the student to read the word (e.g., stir). The peer tutor then praised the student's response and stated the function of the word (e.g., "Stir means you take a spoon and mix.") as nontargeted information. Using a multiple probe across words sets design, the peer tutors taught a total of 15 words to the student using three different cooking products (i.e., hot chocolate mix, muffin mix, microwave popcorn). After the students reached criterion on all three sets of words, the special education teacher assessed generalization by having the students with FMD cook products using novel stimuli (i.e., brands of cooking products).

The peer tutors implemented the project with a high degree of reliability across students: (a) an average of 97% to 99% interobserver reliability agreement and (b) an average of 88% to 97.75% procedural reliability agreement. In addition, the students with FMD were able to read an average of 88% of the target words and define 100% of target words by the final probe session. During the generalization probe session using novel brands of cooking products, the students with FMD were able to read 45% to 90% of the target words and to define 32% to 100% of the target words. In the post-intervention reaction papers, the attitudes of the peer tutors toward students with disabilities showed a positive change, with the peer tutors using more positive terminology (e.g., "students with disabilities" instead of "retarded students") and noting how few differences there were between them.

*Peer delivery of instruction in the general education classroom.* A third study validated the ability of peer tutors to deliver systematic instruction within the general education classroom. Collins, Branson, Hall, and Rankin (2001) sent peer tutors who were part of a peer tutoring class to a secondary composition class with 3 students with FMD. Since the students without disabilities in the composition class would be working on written pieces for their portfolios, the target skill selected for the

students with disabilities was also a writing task. Specifically, the students would learn to write letters that included (a) the date, (b) the greeting, (c) the body, and (d) the closing. Each day before the students with disabilities left for the composition class, one of the special education teachers helped the student fill out a form that stated the following information: (a) the student's name, (b) the date, (c) the recipient of the letter, and (d) the topic of the letter.

Although the original intent of the project was for the general education teacher to conduct the direct instruction and the peer tutor to assist with spelling only, it soon became apparent that there was excess downtime while the students with disabilities waited for the general education teacher to prompt them. Thus, the procedures of the investigation were changed after four sessions to allow the English teacher to deliver the initial task direction to write the letter and the closing feedback statements once the letter was written. A peer tutor assisted the general education teacher by sitting beside each student with FMD during the class and using a system of least prompts (i.e., verbal, model, physical) procedure to help the student write a letter. The peer tutor also collected daily data on the students' independent and prompted responses.

The authors implemented the investigation using a multiple probe across students design. Interobserver reliability agreement was 100% throughout the investigation, and procedural reliability agreement averaged 95%. All students reached criterion on their targeted letter-writing skills, with the average number of sessions to criterion being 14. As in the investigation conducted by Collins et al. (1995), the investigation by Collins, Hedricks, Fetko & Land (2002) also attempted to measure changes in peer attitudes. Again, anonymous reaction papers written by the peers in the composition class showed a positive change following the participation of the first student in the composition class. The investigation failed to measure attitudinal changes toward the subsequent two students with disabilities after they participated in the class.

### Peers as Assistants in Intervention

Peers without disabilities do not have to actually deliver direct instruction to facilitate learning. The following investigations provide examples of how students without disabilities have been involved in systematic strategies to promote generalization and the acquisition of nontargeted information.

*Peer involvement in the special education classroom to facilitate generalization.* Collins, Hall, and Branson (1997) hypothesized that friendships might be facilitated by interactions where peers engaged in activities that were more fun than instructional. To test this hypothesis, the authors involved 21 peers without disabilities in four leisure activities with 3 students with FMD that served as generalization probe sessions. Each day, the special education teacher used a system of least prompts procedure to teach various leisure skills to the students with FMD in the special education classroom. The target skills included the following: (a) passive, independent leisure activity – selecting and watching a television program, (b) passive, leisure activity with others – watching a sports videotape, (c) active, independent leisure activity – playing a computer game, and (d) active leisure activity with others – playing a card game. On Friday of each week, students from an advanced English class who were selected to be “peer buddies” came to the special education classroom to participate in the targeted leisure skills with the students with FMD. These sessions functioned as probes for generalization. The special education teacher instructed the peers to complete the step of the task analysis for the student with FMD if the student exhibited an error while participating in the activities.

Using a multiple probe across activities design replicated across students, Collins et al. (1997) found the system of least prompts procedure delivered by the special education teacher was effective in teaching the leisure skills to the students with FMD. In addition, all of the students with FMD showed an increase in generalization to performing the activities with peer buddies over time. Again, the peers without disabilities showed a positive increase in attitudes toward students with disabilities in

post-intervention reaction papers. Several of the peers also completed written portfolio pieces based on their experiences in the project.

*Peer involvement in the general education setting to facilitate generalization.* Peers without disabilities also facilitated generalization in an investigation by Collins, Hall, Rankin, and Branson (1999). In this investigation, the authors assigned peers from a secondary oral communications class to serve as confederates in a project to teach 4 students with FMD to resist peer pressure. Prior to beginning the project, the authors surveyed peers without disabilities to identify the types of peer pressure that secondary students may have to resist. The resulting categories included (a) behavior that is harmful to one’s health, (b) behavior that is harmful to one’s achievement, and (c) behavior that is harmful to others.

During the project, the students without disabilities attended the oral communications class with the chief purpose of getting to know their peers without disabilities. The authors reasoned that the students with disabilities might be more inclined to submit to peer pressure if they knew the students applying the pressure.

The procedures in the investigation followed those developed by Gast, Collins, Wolery, and Jones (1993) to teach resistance to the lures of strangers. Daily probe sessions took place in integrated school settings (e.g., the hallway between classes). After being told the type of pressure to apply by the special education teacher, a different peer confederate each day approached one of the students with FMD in a predetermined setting. One of the special education teachers loitered nearby and recorded the student’s response when the peer confederate applied the pressure. The special education interrupted the interaction if the students with FMD agreed to perform the inappropriate behavior.

During the initial intervention session, the special education teachers conducted instruction in the special education classroom in which they (a) defined inappropriate behavior, (b) discussed the consequences of this behavior, (c) modeled a correct response to peer pressure (i.e., say “no” and walk away), and (d) led the

students in role playing the correct responses using a constant time delay procedure with a 0-s response interval. On all subsequent days, one of the teachers would interrupt incorrect responses during probe sessions with the peer confederates and immediately conduct an in-vivo practice session of the correct response using a constant time delay procedure. A 5-s response interval was used in the constant time delay procedure. A multiple probe across students design revealed that the students with FMD acquired the target response and generalized it to novel peers, novel settings, and novel types of peer pressure.

*Peer involvement in the general education classroom to deliver nontargeted information.* While it has been established that students with disabilities can learn target behaviors through systematic direct instruction from peers, much of the information peers typically deliver in the general education setting can be classified as nontargeted information. The presentation of nontargeted information does not require a response from the target students; likewise, there is no consequence for the target student responding. The systematic presentation of nontargeted information allows students the opportunity to acquire information that may be useful. It is advantageous because it is a way of inserting additional information into instruction or downtime, thus increasing the efficiency of learning.

Collins et al. (2002) recruited peers without disabilities in a secondary and an elementary general education setting to systematically deliver discrete nontargeted information to included students with FMD. There were 2 target students in each setting, and a student teacher that worked with the students in the general education setting selected the nontargeted information based on input from the students' special education teachers. To extend the results of Werts, Wolery, Holcombe, and Frederick (1993) and Collins, Hall, Branson, and Holder (1999) in comparing the acquisition of related versus unrelated nontargeted information delivered by teachers, Collins et al. (2002) selected nontargeted information for each student that included information that was related to the instructional setting and

information that was not related to the instructional setting.

The general education setting for the secondary students was a Kentucky Studies class; the related information to be presented included three facts about Kentucky (e.g., the name of the capital) and the unrelated information included three safety facts (e.g., how to treat a burn). The general education setting for the elementary students was a theme unit on nutrition; the related information included identification of three pictures of foods, and the unrelated information included letter sounds for one of the students and manual signs of numbers for the other student.

Following baseline conditions, the student teacher began intervention by handing out two different cards each day to various peers in the class that showed the nontargeted information to be presented. She directed the peers to present the nontargeted information to the target students 4-6 times during the class. The peers typically did this during downtimes (e.g., when they went to sharpen a pencil).

The student teacher collected daily probe data within a parallel treatments across sets of facts design replicated across students. This showed her when the students acquired the nontargeted information and whether students acquired related or unrelated facts first. The teacher also collected procedural reliability data once per condition on the peers' delivery of the nontargeted information. The resulting reliability agreement was 100% across peers and condition.

The results showed that most of the students with FMD acquired more unrelated than related facts during the investigation. The authors speculated that the novelty or functionality of nontargeted information may affect its acquisition.

## DISCUSSION

While the investigations described in this review are limited to the work of the author and her colleagues, they address a sample of issues regarding the use of peers without disabilities in the delivery of systematic instruction. The articles suggest that peers

without disabilities can deliver instructional procedures as reliably and effectively as teachers who are trained in special education, and they can provide that instruction reliably and effectively in either the special or the general education setting. In addition, the use of multiple peers as instructors can be an effective way of facilitating generalization, and the use of multiple peers in generalization probe sessions can document that generalization has occurred. Also, students with disabilities appear to learn information from peers that is not targeted for direct instruction when it is presented in a systematic fashion. Finally, when students without disabilities are involved in peer projects with students with mental disabilities, their attitudes toward those students tend to change in a positive manner.

It should be noted that all but one of these studies were conducted with students with functional mental disabilities and their peers in the secondary setting. The results of these studies need to be replicated with students of other age groups and with other types of disabilities. In spite of this limitation, the author offers several guidelines for others who want to involve peers in the delivery of systematic instruction:

1. Teachers should take advantage of peers across settings in a number of ways. While peers can be effective tutors, they also can contribute to learning by participating in generalization activities and by delivering nontargeted information.
2. Peers can be trained to reliably implement systematic instructional procedures and to reliably collect instructional data. Teachers should consider a training package to prepare peers that can include verbal written instructions, information about disabilities, and role-playing.
3. To document fidelity to procedures, teachers periodically should collect procedural and interobserver reliability data. If reliability agreement drops below a set

percentage, peers should be retrained.

4. Peers may take their responsibilities more seriously if they are given incentives to participate. For example, peers can be involved in tutoring activities for course credit, and they can use their experiences as the basis for written English assignments. Involvement in data-based decision-making activities may help peers realize the importance of their involvement.
5. Teachers may want to collect anecdotal data to document changes in peer attitudes over time. This can be a selling point to administrators, general education teachers, and parents in developing peer projects.

## REFERENCES

- Collins, B. C., Branson, T. A., & Hall, M. (1995). Teaching generalized reading of cooking product labels to adolescents with mental disabilities through the use of key words taught by peer tutors. *Education and Training in Mental Retardation and Developmental Disabilities, 30*, 65-75.
- Collins, B. C., Branson, T. A., Hall, M., & Rankin, S. W. (2001). Teaching secondary students with moderate disabilities in an inclusive academic classroom setting. *Journal of Developmental and Physical Disabilities, 13*, 41-59.
- Collins, B. C., Hall, M., & Branson, T. A. (1997). Teaching leisure skills to adolescents with moderate disabilities. *Exceptional Children, 63*, 499-512.
- Collins, B. C., Hall, M., Branson, T., & Holder, M. (1999). Acquisition of related and nonrelated nontargeted information presented by a teacher within an inclusive setting. *Journal of Behavioral Education, 9*, 223-237.
- Collins, B. C., Hall, M., Rankin, S. W., & Branson, T. A. (1999). Just say "No!" and walk away: Teaching students with mental disabilities to resist peer pressure. *Teaching Exceptional Children, 31*, 48-52.
- Collins, B. C., Hendricks, T. B., Fetko, K., & Land, L. (2002). Student-2-student learning in inclusive classrooms. *Teaching Exceptional Children, 34*(4), 56-61.
- Gast, D. L., Collins, B. C., Wolery, M., & Jones, R. (1993). Teaching preschool children with developmental delays to respond to the lures of strangers. *Exceptional Children, 59*, 301-311.
- Kleinert, H. L., Keams, J. F., & Kennedy, S. (1997). Accountability for all students: Kentucky's alternate portfolio assessment for students with moderate and severe cognitive disabilities. *Journal of the Association for Persons with Severe Handicaps, 22*, 88-101.

- Koury, M., & Browder, D. M. (1986). The use of delay to teach sight words by peer tutors classified as moderately mentally retarded. *Education and Training of the Mentally Retarded, 21*, 252-258.
- Miracle, S.A., Collins, B. C., Schuster, J. W., & Grisham-Brown, J. (2001). Peer versus teacher delivered instruction: Effects on acquisition and maintenance. *Education and Training in Mental Retardation and Developmental Disabilities, 36*, 375-385.

Werts, M. G., Wolery, M., Holcombe, A., & Frederick, C. (1993). Effects of instructive feedback related and unrelated to the target behaviors. *Exceptionality, 4*, 81-95.

Wolery, M., Ault, M. J., & Doyle, P. M. (1992). *Teaching students with moderate to severe disabilities: Use of response prompting strategies*. New York, NY: Longman.

**Author's Note:** All correspondence concerning this article should be addressed to Dr. Belva C. Collins, Ed.D., University of Kentucky, Department of Special Education and Rehabilitation Counseling, 229 Taylor Education Building, Lexington, KY 40506-0001 (email: [bcoll01@pop.uky.edu](mailto:bcoll01@pop.uky.edu)).

Table 1

*AN OVERVIEW OF THE STUDIES PRESENTED IN THIS ARTICLE*

Study	Research Focus	Participants	Role of Peers	Strategy	Research Design	Results
Miracle et al. (2001)	Peer- vs. teacher-delivered instruction	4 students with FMD, 5 peer tutors, and special education teacher in secondary setting	Deliver direct instruction in special education classroom	Constant time delay to teach sight words	Alternating treatments	Peer-delivered instruction as effective as teacher-delivered instruction

---

Collins et al. (1995)	Key words using multiple exemplars	4 students with FMD and 26 peer tutors in secondary setting	Deliver direct instruction in special education classroom	Constant time delay to teach cooking words with multiple exemplars and non- targeted informa- tion	Multiple probes across word sets replicated across students	Key words and non- targeted informa- tion presented by peer tutors acquired and general- ized
--------------------------	---	--	--	--	---	--

---

---

Collins et al. (2001)	Letter writing in inclusive setting	3 students with FMD, English teacher, and variety of peer tutors in secondary setting	Deliver prompts and provide assistance with spelling in general education class	System of least prompts to teach letter- writing skills	Multiple probe across students	English teacher need peer tutor support to deliver instruction in inclusive setting
--------------------------	--	---	---	--	---	---

---



---

Collins et al. (1997)	Generalization of leisure skills	3 students with FMD, special education teacher, and 21 peer buddies in secondary setting	Conduct generalization probes as “peer buddies” in special education class	System of least prompts to teach leisure skills	Multiple probe across skills replicated across students	Students acquired and generalized leisure skills
-----------------------	----------------------------------	--	--	---	---	--

---

---

Collins et al. (1999)	Generalization of safe response	3 students with FMD, special education teachers, and variety of peers in secondary setting	Act as confederates during in vivo probe and training trials in school setting	Constant time delay with multiple exemplars to teach resistance to peer pressure	Multiple probe across students	Students acquired and generalized resistance response
-----------------------	---------------------------------	--	--	--	--------------------------------	---

---

---

Collins et al. (2002)	Acquisi- tion of related and unrelated non- targeted informa- tion from peers in inclusive setting	2 students with FMD and their peers in secondary setting; 2 students with FMD and their peers in elemen- tary setting	Deliver non- targeted informa- tion in inclusive general education class	Syste- matically deliver non- targeted informa- tion without a specific instruc- tional strategy	Parallel treatments designs replicated across 2 experi- ments	Students acquired unrelated non- targeted informa- tion than related non- targeted informa- tion
--------------------------	--	--	--	---	---	---

---

## HUMAN TIME-BASED SCHEDULE PERFORMANCE: EXPLORING INTERACTIONS BETWEEN INSTRUCTIONS AND RESPONSE-REINFORCER CONTIGUITY

Mickey Keenan  
University of Ulster at Coleraine  
Ken P. Kerr  
Saplings Model of Education  
County Kildare, S. Ireland

Spontaneous regulation of human behavior on a simple time-based mixed schedule of reinforcement was studied as a function of instructions and response-reinforcer contiguity. Seventeen subjects were exposed to a mixed fixed-interval (FI) 10 s schedule (Condition A) and a conjunctive fixed-time 10 s fixed ratio 1 (Condition B) schedule of reinforcement using either an ABAB or a BABA design across 5 sessions. One of three instructional categories was employed (normal, ratio, and interval instructions) on a simple button pressing task. Subjects either produced a high rate of responding with short postreinforcement pauses (PRPs), a high rate of responding decreasing to a low rate with long PRPs, or a low rate of responding with long PRPs. Evidence of spontaneous regulation in terms of consistent differential responding across components was displayed only under interval instructions. These results highlight the interaction of instructions with the role of response-reinforcer contiguity in controlling spontaneous regulation of human performance on simple time-based schedules of reinforcement.

---

Typical human performance on a fixed-interval (FI) schedule of reinforcement is commonly described as either a high rate performance with a small postreinforcement pause (PRP), or a low rate performance with a long PRP (e.g., Leander, Lippman & Meyer, 1968; Lippman & Meyer, 1967; Weiner, 1969). These findings differ from nonhuman performance on the FI where a scallop or break-and-run is the typical pattern (Baron & Leinenweber, 1994; Hyten & Madden, 1993; Perone, Galizio & Baron, 1988). Attempts to explain the discrepant findings between human and nonhumans, and the differences in findings between humans, have led to the development of a number of research areas. These include studies concerned with the effect of the experimental setting (Barnes & Keenan, 1989; Orne, 1962), the effect of conditioning histories (e.g., Weiner, 1969, 1972), and the determinants of choice (e.g., Belke, Pierce, & Powell, 1989; Darcheville, Riverère, & Wearden, 1993).

One of the defining features of human behavior, the complexity of the verbal repertoire, also has drawn special attention in the study of schedules of reinforcement. Many studies have investigated verbal behavior by focusing on the role of either experimenter-generated or self-generated instructions on schedules of reinforcement (e.g., Baron & Galizio, 1983; Galizio, 1979; Hayes,

Brownstein, Zettle, Rosenfarb, & Korn, 1986; LeFrangois, Chase, & Joyce, 1988; Matthews, Shirnoff, Catania, & Sagvolden, 1977). Conclusions to date highlight the power of instructions in generating responding which may be insensitive to the programmed contingencies (Baron & Galizio; for a more recent analysis see Madden, Chase & Joyce, 1998). Some researchers have also pointed to the role of self-generated instructions in determining behavior (Bentall & Lowe, 1987; Lowe, 1979).

Before we can fully understand the role played by verbal behavior in human schedule performance we must have a clear understanding of how the formal structure of a schedule of reinforcement critically determines the dynamics involved in behavioral adaptation. To date most research investigating the interaction between instructions and programmed contingencies have employed schedules simply as tools for generating behavior. The problem with this is that we do not fully understand how the tools themselves function. That is, we are limited in the extent to which we understand schedules as contexts for controlling dynamic processes in the multiple determination of behavior (Keenan & Toal, 1991). If we are limited in our understanding of these fundamental issues, the addition of another layer of complexity (verbal behavior) makes our task more difficult. A detailed analysis of the workings of simple

schedules of reinforcement is therefore necessary. Some authors, however, have suggested that even so-called simple schedules of reinforcement are too complex and that the analysis of independent variables brought into play will lead ultimately to failure (Davey, 1987; Zeiler, 1984).

Although the task of identifying the controlling variables in a schedule of reinforcement is complicated, it is by no means impossible. For example, in the nonhuman experimental paradigm Keenan and Leslie (1984, 1986) showed that four interacting variables play a crucial role in determining baseline performance on the FI schedule of reinforcement. The variables noted were (a) the fixed time (FT) component in which a fixed period of time must elapse before reinforcement is made available; (b) the single response contingency (FR 1) where one response after the FT component has elapsed is necessary to produce reinforcement; (c) the time from the reinforcer presentation to the location of the next response- dependency; and (d) response-reinforcer contiguity where the first response after the FT component has elapsed produces immediate reinforcer delivery.

Keenan and Leslie (1986) noted that there were three schedules composed of similar constituent elements, namely the FI schedule, the conjunctive FT FR 1, and the recycling conjunctive FT FR 1. Despite the regular temporal placement of reinforcement on these schedules, their differing structures produce different performances. Critical to these performances is the role played by response-reinforcer contiguity. Of these schedules, only the FI guarantees response-reinforcer contiguity on the completion of the FT component. On the conjunctive schedule contiguity between the response and reinforcer delivery may or may not occur depending on the temporal placement of the response. If a single response occurs, contiguity depends on whether this occurs inside the FT component. If responding occurs only after the FT component has elapsed, response-reinforcer contiguity is guaranteed. However, if a single response satisfying the FR 1 criterion occurs within the FT component a delay occurs

between the response and the reinforcer delivery. The recycling conjunctive FT FR 1 also allows a delay to occur between responding and reinforcer delivery but the probability of response-reinforcer contiguity is reduced. If responding does not occur inside the FT component the reinforcer is withheld and the next FT component begins. If responding does occur inside the FT component a delay may be encountered between the response and reinforcer delivery. If a response occurs at the end of the FT component fortuitous contiguity may occur.

The powerful effects of contiguity on both response rate and patterning were observed when Keenan and Leslie altered a recycling conjunctive schedule to increase the incidence of response-reinforcer contiguity. Under the normal recycling conjunctive FT 30 s FR 1 rats showed a pause-respond-pause pattern with responding occurring in the middle of the interreinforcer interval. After introducing the programmed response-reinforcer contiguity PRP's increased, the overall response rate increased, and the response pattern changed to resemble an FI scallop. A more recent study using rats by Keenan (2000) systematically controlled the frequency of response-reinforcer contiguity on a schedule with periodic reinforcer presentations. It was found again that both response patterning and response rates were affected only this time performances varied quite dramatically across subjects, some even resembling human performance.

The general approach of dissecting and rebuilding time-based schedules has been extended to human schedule performance by Barnes and Keenan (1994). In considering human differences in response rate and patterning, they exposed human subjects to one session consisting of 75 trials in total on a FI 10 s schedule and a conjunctive FT 10 s FR 1 schedule of reinforcement. Subjects were assigned to either a minimal instruction or a ratio instruction group with equal numbers assigned to one of four conditions; (a) exposure to 25 trials on the FI schedule followed by 50 trials on the conjunctive schedule; (b) exposure to 50 trials on the FI followed by 25 trials on the conjunctive schedule; (c) exposure to 25 trials on the conjunctive schedule followed by 50

trials on the FI schedule and (d) exposure to 50 trials on the conjunctive schedule followed by 25 trials on the FI schedule. In each session the transition across schedules was unsignalled. Findings suggested that response-reinforcer contiguity, as guaranteed on the FI schedule, helped to generate high response rates. Differential response rates and patterning also were observed across the schedules. Another important finding was that the ratio-based instructions appeared to facilitate the response generating function of response-reinforcer contiguity. That is, under ratio instructions they found higher rates of responding on the FI schedule than on the conjunctive schedule.

The focus on the interactive effects of instructions with the constituent elements of the FI and related time-based schedules of reinforcement marks a shift away from the pure contingency analysis of nonhuman behavior by Keenan and Leslie (1986). However, results from Barnes and Keenan (1994) are at best tentative due to several methodological shortcomings. Most important of these was the use of only one session, with only one transition across the schedules. Previous studies on human FI performance suffer from other interpretative problems when either a single FI schedule (e.g., Buskist, Bennett, & Miller, 1981; Lippman & Meyer, 1967) or a multiple schedule of reinforcement (e.g., Baron, Kaufinan, & Stauber, 1969; Galizio, 1979) have been used. When only a single schedule is employed, an accurate assessment of the control exerted by the schedule's dynamic independent variables becomes difficult (cf. Baron & Galizio, 1983). However, by employing two distinct schedules created from the same basic components it is possible to distinguish the way in which the differences in the formal structure of the schedules determines the dynamic properties of behavior on those schedules. The problem with using a multiple schedule of reinforcement is that the control exerted by the dynamic independent variables becomes confounded with the discriminative properties acquired by the stimuli associated with each component schedule. Furthermore, differences in the rate and temporal distribution of reinforcement across components in a multiple schedule may also contribute to differences in responding

between components (see Catania, 1970). The use of a mixed schedule in the current research avoids these interpretative problems.

With these considerations in mind, the current study continues the analysis of the co-determining effects of instructions and response-reinforcer contiguity begun by Barnes and Keenan (1994). Using three instructional categories (ratio-based, interval-based, and normal instructions) over a number of sessions, we examined the response generating effect of response-reinforcer contiguity on a mixed FI 10 s, conjunctive FT 10 s FR 1 schedule of reinforcement with an increased number of transitions per session.

## METHOD

### Subjects

Seventeen undergraduate students enrolled at the University of Ulster, participated in the study. Each subject was paid a flat fee of £1.50 per session.

### Apparatus

Subjects were seated in a small experimental room in front of a British Broadcasting Corporation (BBC) Model B Microcomputer with a Cumana (model 5400) floppy disk drive and a Kaga Denshi (model KG-12NB-N) computer monitor. The experimental screen consisted of a 7 x 7 square grid in which a box moved from left to right each time the schedule criteria was met. The label "Start" was displayed at the bottom left of the grid with "Finish" displayed at the top right. On completion of the first row of 7 squares the box reset to the first square in the second row. Whilst the goal of the experiment was to move the box from the Start point to the Finish point the experiment was deemed finished when the specified number of trials was met.

### Schedules of Reinforcement

A FI 10 s (Condition A) and a conjunctive FT 10 s FR1 (Condition B) schedule of reinforcement alternated for some subjects on an ABAB design and alternated for other subjects on a BABA design. In the FI 10 s

schedule any response within the 10-s interreinforcer interval had no programmed consequence; the first response after the 10-s interval had elapsed resulted in reinforcer delivery. On the conjunctive FT 10 s FRI schedule at least one response anywhere within the 10-s interval produced a reinforcer at the end of that interval. A response after the 10-s interval elapsed produced immediate reinforcer delivery if responding had not occurred earlier. In any one session each subject was exposed to 30 interreinforcer intervals on each schedule during each condition (120 interreinforcer intervals per session).

### Procedure

Subjects were randomly assigned to one of three instructional categories, normal, ratio, or interval instructions. Normal instructions were as follows:

You can move the box through the square grid from start to finish by pressing the spacebar. Press the spacebar to begin the session. Do not leave until the computer tells you the session is over.

Ratio instructions were similar except that the end of the first sentence read:

by pressing the spacebar a random number of times.

Interval instructions were similar with the end of the first sentence reading:

by pressing the spacebar at various intervals.

If a subject asked any questions, the relevant instructions were repeated.

### RESULTS

Results presented here for each subject include the number of responses within each interreinforcer interval and postreinforcement pause (PRP) durations across successive conditions.

### Normal Instructions

Figure 1 and Table 1 show the performances for Jean and Sharon on the ABAB design where A represents the F1 10 s schedule and B represents the conjunctive FT 10 s FR 1 schedule. Jean responded at a high rate with extremely short PRPs across Sessions 1-3. Half way through Condition A in Session 4 there was a marked decrease in responding with a corresponding increase in the PRP duration, though they were still less than 1 s. This performance persisted across the remainder of the session and again in Session 5 although there were a few intervals with increased responding early in this session.

Sharon responded at a high rate in Session 1, with the exception of the second occurrence of Condition A where responding decreased to a low rate. PRP durations in Session 1 were also extremely short and variable until the second occurrence of Condition A where extended PRPs were recorded. A low response rate with mainly extended PRPs across all conditions was recorded in Sessions 2-5. As with Jean there was no evidence of differential responding across Conditions A and B.

Figure 2 and Table 2 show performances for Miriam, Margaret and Clare on the BABA design. Miriam responded at a low rate with extended PRPs throughout most of the experiment. On three occasions, the last condition of Session 2 and first half of Session 3 there was a slight increase in responding. Margaret responded at a low rate in Session 1 with extended PRPs, a pattern that persisted across all subsequent sessions. Clare responded at a high rate with extremely short PRPs in Session 1. A low rate of responding with extended PRPs was recorded in Session 2. This pattern of responding persisted across the remaining sessions.

### Ratio Instructions

Figure 3 and Table 3 show performances for David, Brendan, and Jason on the ABAB design. David responded with a consistent high rate and extremely short PRPs across Sessions 1 to 5. Brendan responded similarly in Sessions 1-3. From Session 4

onwards, however, there was a substantial decrease in response rate with extended PRPs. Jason showed an overall low rate of responding with relatively large PRPs across sessions.

Figure 4 and Table 4 show performances for Carty, Brofie, and Alan on the BABA design. In Session 1 responding for Carty began initially at a high rate but then dropped to a substantially low rate thereafter with extended PRPs. Alan also began Session 1 by responding at a relatively high rate with generally short PRPs. However, by the final Condition A on the first session responding decreased to a low rate with a substantial increase in PRP durations. This pattern persisted across the remaining sessions. Brofie responded at a low rate across all conditions and sessions. However, in Sessions 2 & 3 there was some evidence of differential control over PRP durations by Conditions A & B with pausing shortest in Condition B (Conjunctive schedule). Across the remaining sessions short PRP durations predominated in both conditions.

#### Interval instructions

Figure 5 and Table 5 show results for Patricia, Kevin, and Rosemary on the ABAB design. Responding for Patricia was initially at a low rate but this increased slightly across Session 1. The increase in responding was maintained across the remaining sessions. Apart from the first two conditions on the last session the PRP durations were relatively large across sessions. Kevin responded at a low rate across all sessions. PRPs in Session 1 were mainly extended with a decrease occurring on exposure to the final conjunctive schedule. In the Sessions 2-3 there was some evidence of differential patterning in PRP duration with shorter PRPs occurring in Condition B. Rosemary responded at a low rate with extended PRPs across all sessions.

Figure 6 and Table 6 show the performance of Paul, Ciaran, and J.P. on the ABAB design. Although Paul responded at a very low rate across sessions there was evidence of differential responding across conditions with more responding occurring in intervals on the conjunctive schedule (Condition B) than on the

FI schedule (Condition A) (Sessions 2-5). Differential responding was also evident in PRP durations with longer PRPs recorded in intervals on the FI schedules. Ciaran responded at a low rate across all sessions. Differential responding across conditions was evident with response rates being generally higher in Condition B. Across Session 1 PRP durations increased and by Sessions 2-5 there were differential effects across conditions with lower PRP durations observed in Condition B. J.P. produced a high rate of responding with mainly short PRPs across all sessions.

#### DISCUSSION

Subjects were assigned either to an ABAB or BABA sequence of conditions where A and B represent an FI 10 s and a conjunctive FT 10 s FR 1 schedule of reinforcement respectively. Each sequence of conditions had three instructional categories; normal, interval, or ratio. Across these conditions four general response patterns were observed: (a) a consistent high rate across sessions, (b) a high rate decreasing to a low rate across sessions, (c) a consistent low rate across sessions, and (d) differential responding across components. The initial three patterns are typical of performance on the FI schedule of reinforcement (cf. Baron & Galizio, 1983).

Responding under normal instructions in both the ABAB and BABA sequence of conditions showed two general types of performances; (1) a high rate of responding with short PRPs and (2) a low rate of responding with extended PRPs. On two occasions the high rate of responding changed to a low rate during exposure to the FI schedule. No consistent evidence of systematic change in response rate or patterning was observed across the different conditions.

Performance under ratio instructions was similar to that under normal instructions. One subject in the ABAB sequence of conditions showed a change in responding from a high to low rate. This change occurred in the FI schedule where a delay between responding and reinforcer delivery is not possible. Brofie showed evidence of differential response rates and PRP durations in the BABA sequence of



conditions, although this was not maintained across all sessions. Response rates generally in the ABAB sequence of conditions were higher than response rates in the BABA sequence of conditions.

Performance under interval instructions showed more evidence of differential control of response rate and PRP duration. In the ABAB sequence of conditions, Kevin showed differential PRP durations across conditions (Session 2 & 3). In the BABA sequence of conditions Paul responded differentially across conditions in Sessions 1-5 and Ciaran responded differentially across conditions in Sessions 2-4.

Regarding the occurrence of differential responding across components, there was no systematic relation to instructional categories. That differential responding occurred at all, though, is testimony to the potential usefulness of this experimental strategy of using this combination of schedule components. Since the overall rate of reinforcement on each schedule was largely determined by the same FT duration, any differential responding presumably indicates the sensitivity of responding to the differing dynamics afforded by the different structures of the two schedules. In particular, the conjunctive permits delays between responding and reinforcer delivery but the FI schedule does not.

The three other response patterns observed (consistent high rate, consistent low rate, and high rate decreasing to a low rate) suggest that responding might be insensitive to programmed contingencies. However, based on the data from 6 people under ratio instructions, an alternative analysis may be that competing contingencies vie for control over responding (Hayes & Wolf, 1984). In this case, a possible candidate is the operation of a sequencing effect. For example, under ratio instructions 2 out of 3 people produced typical high rate responding across the ABAB sequence of conditions. This compares with Leander, Lippinan, and Meyer (1968) and Barnes and Keenan (1994). Exposure to the initial FI schedule under ratio instructions resulted in a high rate performance which was maintained across the sequence of conditions, regardless of any disruption of response-reinforcer contiguity. For the three

people on the BABA sequence of conditions, under ratio instructions, this high rate pattern was not apparent. It is possible that the initial delays between responding and reinforcer delivery in Condition B may exert a greater effect when there has been no previous history of response-reinforcer contiguity on the FI schedule. The effect of the delay may also be rate specific. That is, the delay may only exert an effect on responding whenever a low response rate is observed. Overall, it appears that any effect of disrupting response-reinforcer contiguity may be dependent to some degree on the sequencing of conditions and the response rate under certain instructional categories.

To conclude, although the effect of response-reinforcer contiguity in determining response rate and patterning has been shown clearly in nonhuman research (Keenan, 2000; Lattal & Ziegler, 1982) the effect is not as clear with humans. A general point about human schedule performance research is worth considering though. Advances in behavior analysis generally hinge on the ability to control behavior by the manipulation of appropriate contingencies of reinforcement. Within the debate on sensitivity to contingencies (Madden, Chase, & Joyce, 1998) there is perhaps a lesson to be had if we look at things from an applied perspective. When insensitivity to contingencies is observed, differential responding might best be considered as a target behavior that we have difficulty in producing. Viewed in this way, the experimental task at hand is to either conduct a rigorous assessment of the effectiveness of the reinforcer employed (e.g., Baer, Tishelman, Degler, Osnes, & Stokes, 1992; Smith, Iwata, & Shore, 1995) or to adjust the contingencies so as to produce the required target behavior. On the flip side, it would prove interesting to determine if it was possible to identify features of a schedule that could be changed to produce 'insensitivity' to contingencies for nonhumans. In the context of the search for understanding basic schedule effects these suggestions stem from a conceptualisation of a schedule that differs from its as more usual depiction as a device for arranging a relation between a response and a reinforcer. Keenan and Toal (1991) put it this way:

"... a schedule is more properly conceived as providing an opportunity for observing the dynamic behavioral system that "crystallizes out" when a biological system is exposed to environmental constraints... At any one instance, the characteristics of the behavioral system are dependent upon the interplay between the "plasticity" or dynamic limitations inherent in the adaptiveness of the biological system, and the dynamics imposed across time by the structure of the prevailing contingencies." (p. 113)

## REFERENCES

- Baer, R. A., Tishelman, A. C., Degler, J. D., Osnes, P. S., & Stokes, T. F. (1992). Effects of self-vs. experimenter- selection of rewards on classroom behavior in young children. *Education and Treatment of Children, 15*, 1-14.
- Bames, D., & Keenan, M. (1994). Response-reinforcer contiguity and human performance on simple time-based reinforcement schedules. *The Psychological Record, 44*, 63-90.
- Bames, D., & Keenan, M. (1989). Instructed human fixed -interval performance: The effects of the experimental setting. *The Psychological Record, 39*, 351-363.
- Baron, A., & Galizio, M. (1983). Instructional control of human operant behaviour. *The Psychological Record, 33*, 495-520.
- Baron A., Kaufinan, A., & Stauber, K. A. (1969). Effects of instructions and reinforcement-feedback on human operant behavior maintained by fixed-interval reinforcement. *Journal of Experimental Analysis of Behavior, 12*, 701-712.
- Baron, A., & Leinenweber, A. (1994). Molecular and molar analyses of fixed-interval performance. *Journal of the Experimental Analysis of Behavior, 61*, 11-18.
- Belke, T. W., Pierce, W. D., & Powell, R. A. (1989). Determinants of choice for pigeons and humans on concurrent-chain schedules of reinforcement. *Journal of the Experimental Analysis of Behavior, 52*, 97-110.
- Bentall, R. P., & Lowe, C. F. (1987). The role of verbal behavior in human learning. III: Instructional effects in children. *Journal of Experimental Analysis of Behavior, 47*, 177-190.
- Buskist, W. F., Jr., & Bennett, R. H., Miller, H. L. (1981). Effects of instructional constraints on human fixed-interval performance. *Journal of Experimental Analysis of Behavior, 35*, 217-225.
- Catania, A. C. (1970). Reinforcement schedules and psychophysical judgements: A study of some temporal properties of behavior. In W. N. Schoenfeld (Ed.), *The theory of reinforcement schedules* (pp. 1-42). New York: Appleton-Century Crofts.
- Darcheville, J. C., Riverère, V., & Wearden, J. H. (1993). Fixed-interval performance and self-control in infants. *Journal of the Experimental Analysis of Behavior, 60*, 239-254.
- Davey, G. C. L. (1987). *Animal learning and conditioning*. London: Macmillan Education.
- Galizio, M. (1979). Contingency-shaped and rule-governed behavior: Instructional control of human loss avoidance. *Journal of the Experimental Analysis of Behavior, 31*, 53-70.
- Hayes, S. C., Brownstein, A. L., Zettle, R. D., Rosenfarb, L., & Kom, Z. (1986). Rule-governed behavior and sensitivity to changing consequences of responding. *Journal of the Experimental Analysis of Behavior, 45*, 237-256.
- Hayes, S. C., & Wolf, M. R. (1984). Cues, consequences, and therapeutic talk: Effect of social context and coping statements on pain. *Behavior Research and Therapy, 22*, 385-392.
- Hyten, C., & Madden, G. J. (1993). The scallop in human fixed-interval research: A review of problems with data description. *The Psychological Record, 43*, 471-500.
- Keenan, M. (2000). Periodic response-reinforcer contiguity: Temporal control but not as we know it! *The Psychological Record, 49*, 273-297.
- Keenan, M., & Leslie, J. C. (1984). Separating response dependency and response-reinforcer contiguity in a recycling conjunctive schedule. *Journal of the Experimental Analysis of Behavior, 41*, 203-210.
- Keenan, M., & Leslie, J. C. (1986). Varying response-reinforcer contiguity in a recycling conjunctive schedule. *Journal of the Experimental Analysis of Behavior, 45*, 317-332.
- Keenan, M., & Toal, L. (1991). Periodic reinforcement and second-order schedules. *The Psychological Record, 41*, 87-115.
- Lattal, K. A., & Ziegler, D. R. (1982). Briefly delayed reinforcement: An interresponse time analysis. *Journal of the Experimental Analysis of Behavior, 37*, 407-416.
- Leander, J. D., Lippman, L. G., & Meyer, M. E. (1968). Fixed-interval performance as related to subjects' verbalization of the reinforcement contingency. *The Psychological Record, 18*, 469-474.
- LeFrangois, J. R., Chase, P. N., & Joyce, J. H. (1988). The effects of a variety of instructions on human fixed-interval performance. *Journal of the Experimental Analysis of Behavior, 49*, 383-393.
- Lippman, L. G., & Meyer, M. E. (1967). Fixed-interval performance as related to instructions and to subjects' verbalization of the reinforcement contingency. *Psychonomic Science, 8*, 135-136.
- Lowe, C. F. (1979). Determinants of human operant behavior. In M. D. Zeiler & P. Harzem (Eds.), *Reinforcement and the structure of behavior* (pp. 159-192). New York: Wiley.
- Madden, J., Chase, P. N., & Joyce, J. H. (1998). Making sense of sensitivity in the human operant literature. *The Behavior Analyst, 21*, 1-12.
- Matthews, B. A., Shimoff, E., Catania, A. C., & Sagvolden, T. (1977). Uninstructed responding: Sensitivity to ratio and interval contingencies. *Journal of the Experimental Analysis of Behavior, 27*, 453-467.
- Ome, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to the demand characteristics and their implications. *American Psychologist, 19*, 776-783.
- Perone, M., Galizio, M., & Baron, A. (1988). The relevance of animal-based principles in the laboratory study of human

- operant conditioning. In G. C. L. Davey & C. Cullen (Eds.), (1988). *Human operant conditioning and behaviour modification*. Bury St. Edmunds: St. Edmundsbury Press.
- Smith, R. G., Iwata, B. A., & Shore, B. A. (1995). Effects of subject-versus experimenter- selected reinforcers on the behavior of individuals with profound developmental difficulties. *Journal of Applied Behavior Analysis*, 28, 61-72.
- Weiner, H. (1969). Controlling human fixed-interval performance. *Journal of the Experimental Analysis of Behavior*, 12, 349-373.
- Weiner, H. (1972). Controlling human fixed-interval performance with fixed-ratio responding or differential reinforcement-of-low -rate responding in mixed schedules. *Psychonomic Science*, 25, 191-192.
- Zeiler, M. D. (1984). The sleeping giant: Reinforcement schedules. *Journal of the Experimental Analysis of Behavior*, 42, 485-493.

**Author's Note:** This research was conducted while Ken Kerr was holder of a Dept. of Education for N. Ireland (DENI) doctoral grant under the supervision of Mickey Keenan. Reprints may be obtained from Mickey Keenan at the School of Psychology, University of Ulster, at Coleraine, Cromore Road, Coleraine, County L' Derry, N. Ireland, BT 52 ISA, email: [m.keenan@ulst.ac.uk](mailto:m.keenan@ulst.ac.uk), or from Ken Kerr at Saplings Model of Education, Wheelam, Milltown, Newbridge, County Kildare, S. Ireland.

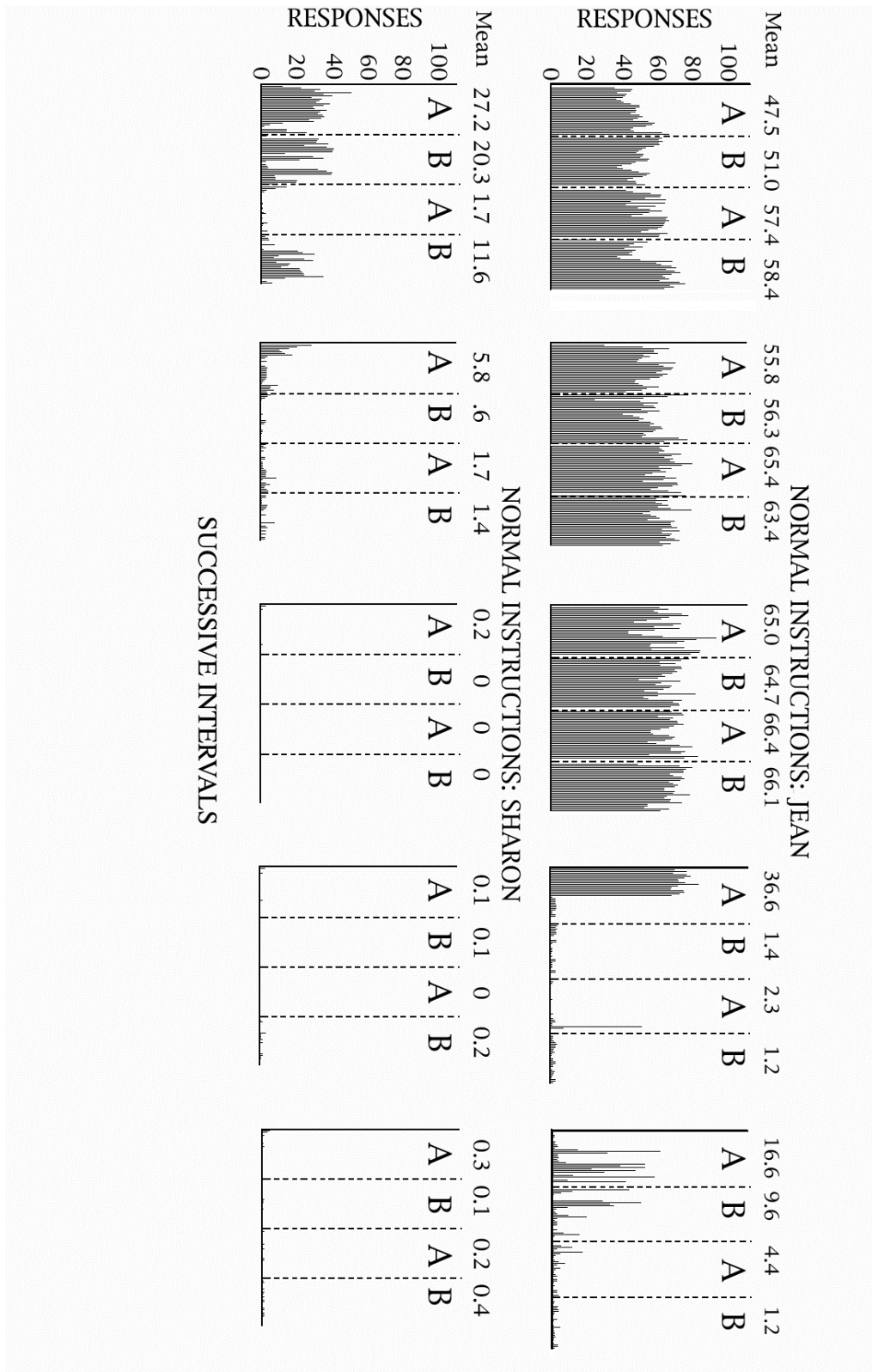


Figure 1. Number of responses during each FT component of Conditions A & B in each session for Jean and Sharon.

Table 1

Normal Instructions: ABAB

PRP durations in seconds for each subject across conditions and across sessions

	A		B		A		B	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
Jean	<b>0.13</b>	0.14	<b>0.10</b>	0.16	<b>0.01</b>	0.13	<b>0.08</b>	0.16
	<b>0.12</b>	0.31	<b>0.12</b>	0.16	<b>0.07</b>	0.09	<b>0.14</b>	0.22
	<b>0.11</b>	0.10	<b>0.09</b>	0.12	<b>0.10</b>	0.11	<b>0.08</b>	0.10
	<b>0.30</b>	0.21	<b>0.66</b>	0.74	<b>1.01</b>	0.95	<b>0.75</b>	0.73
	<b>0.64</b>	0.55	<b>0.67</b>	0.57	<b>0.86</b>	0.82	<b>0.84</b>	0.88
Sharon	<b>0.18</b>	1.06	<b>0.16</b>	3.05	<b>10.52</b>	9.21	<b>1.0</b>	3.75
	<b>6.18</b>	4.78	<b>10.53</b>	10.06	<b>8.31</b>	7.10	<b>8.97</b>	9.54
	<b>10.77</b>	10.74	<b>10.96</b>	11.09	<b>10.87</b>	11.35	<b>10.84</b>	10.73
	<b>10.99</b>	10.89	<b>10.90</b>	10.80	<b>11.30</b>	11.69	<b>10.84</b>	10.67
	<b>10.91</b>	10.72	<b>10.69</b>	10.64	<b>10.69</b>	10.60	<b>10.23</b>	10.34



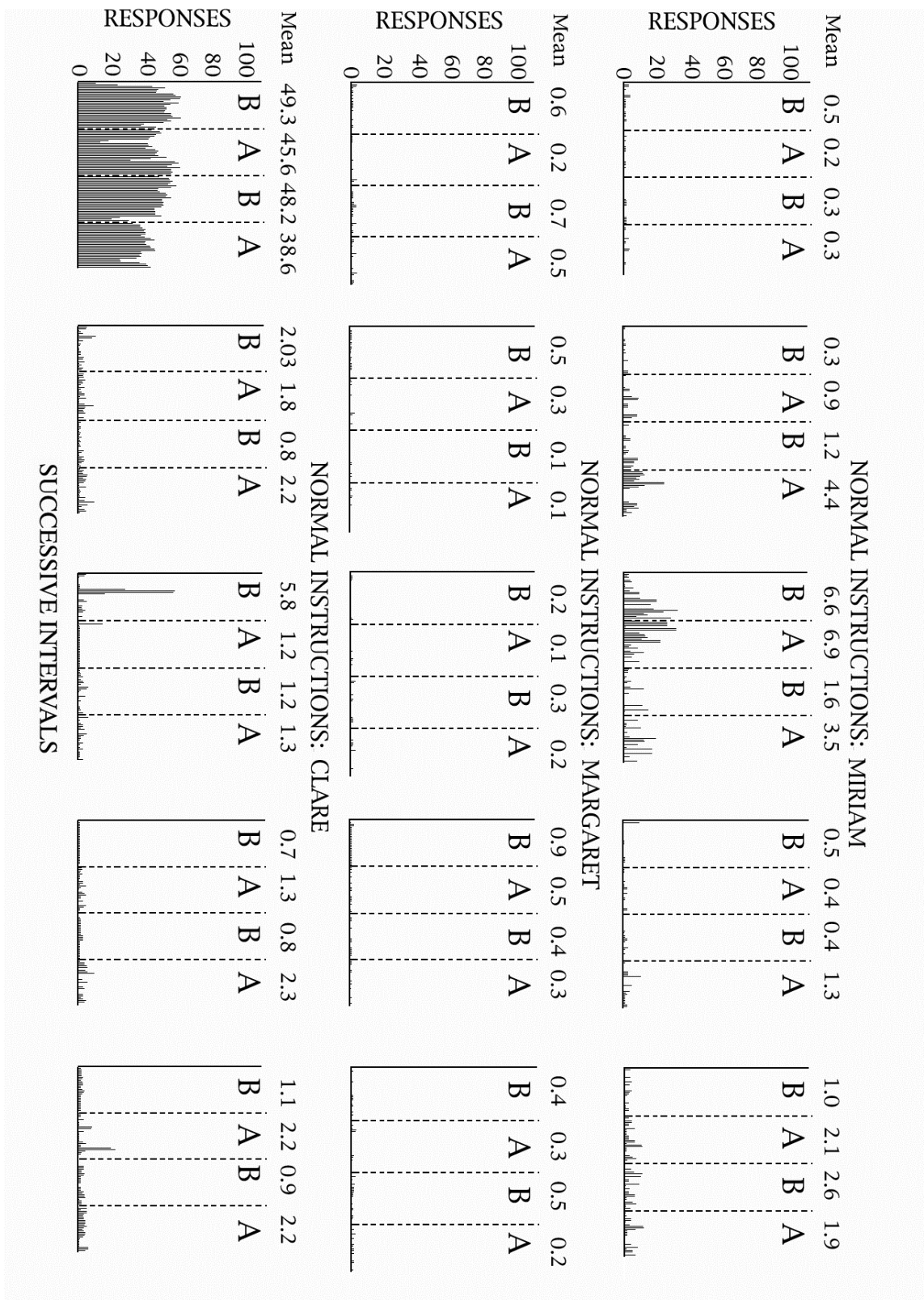


Figure 2. Number of responses during each FT component of Conditions A & B in each session for Miriam, Margaret, and Claire.

Table 2

Normal Instructions: BABA

PRP durations in seconds for each subject across conditions and across sessions

	B		A		B		A	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
Miriam	<b>11.12</b>	11.61	<b>12.76</b>	12.67	<b>12.43</b>	12.97	<b>12.05</b>	12.25
	<b>13.23</b>	12.31	<b>11.90</b>	11.48	<b>10.21</b>	9.80	<b>8.66</b>	8.05
	<b>8.08</b>	9.19	<b>8.52</b>	10.74	<b>0.36</b>	12.57	<b>0.75</b>	10.88
	<b>13.67</b>	13.04	<b>14.56</b>	13.56	<b>13.59</b>	13.43	<b>9.34</b>	9.60
	<b>10.74</b>	10.57	<b>8.36</b>	9.14	<b>8.68</b>	8.71	<b>9.72</b>	10.88
Margaret	<b>9.85</b>	8.47	<b>17.73</b>	15.36	<b>9.86</b>	9.68	<b>10.10</b>	9.93
	<b>9.87</b>	10.2	<b>13.62</b>	12.33	<b>13.71</b>	13.35	<b>13.78</b>	13.14
	<b>13.55</b>	12.58	<b>13.35</b>	12.62	<b>12.40</b>	12.12	<b>13.26</b>	12.80
	<b>9.76</b>	9.82	<b>10.04</b>	10.04	<b>10.10</b>	10.25	<b>10.14</b>	10.35
	<b>10.29</b>	10.26	<b>10.43</b>	10.24	<b>101.2</b>	10.25	<b>10.19</b>	10.27
Clare	<b>0.14</b>	0.42	<b>0.01</b>	0.16	<b>0.20</b>	0.27	<b>0.13</b>	0.21
	<b>9.62</b>	8.69	<b>9.13</b>	8.21	<b>9.80</b>	9.20	<b>7.80</b>	8.15
	<b>8.79</b>	7.40	<b>7.06</b>	6.42	<b>8.61</b>	7.68	<b>1.96</b>	4.67
	<b>1.09</b>	6.46	<b>7.38</b>	6.70	<b>0.70</b>	4.74	<b>9.03</b>	7.33
	<b>0.67</b>	2.58	<b>10.14</b>	1.83	<b>9.63</b>	9.54	<b>9.46</b>	8.87

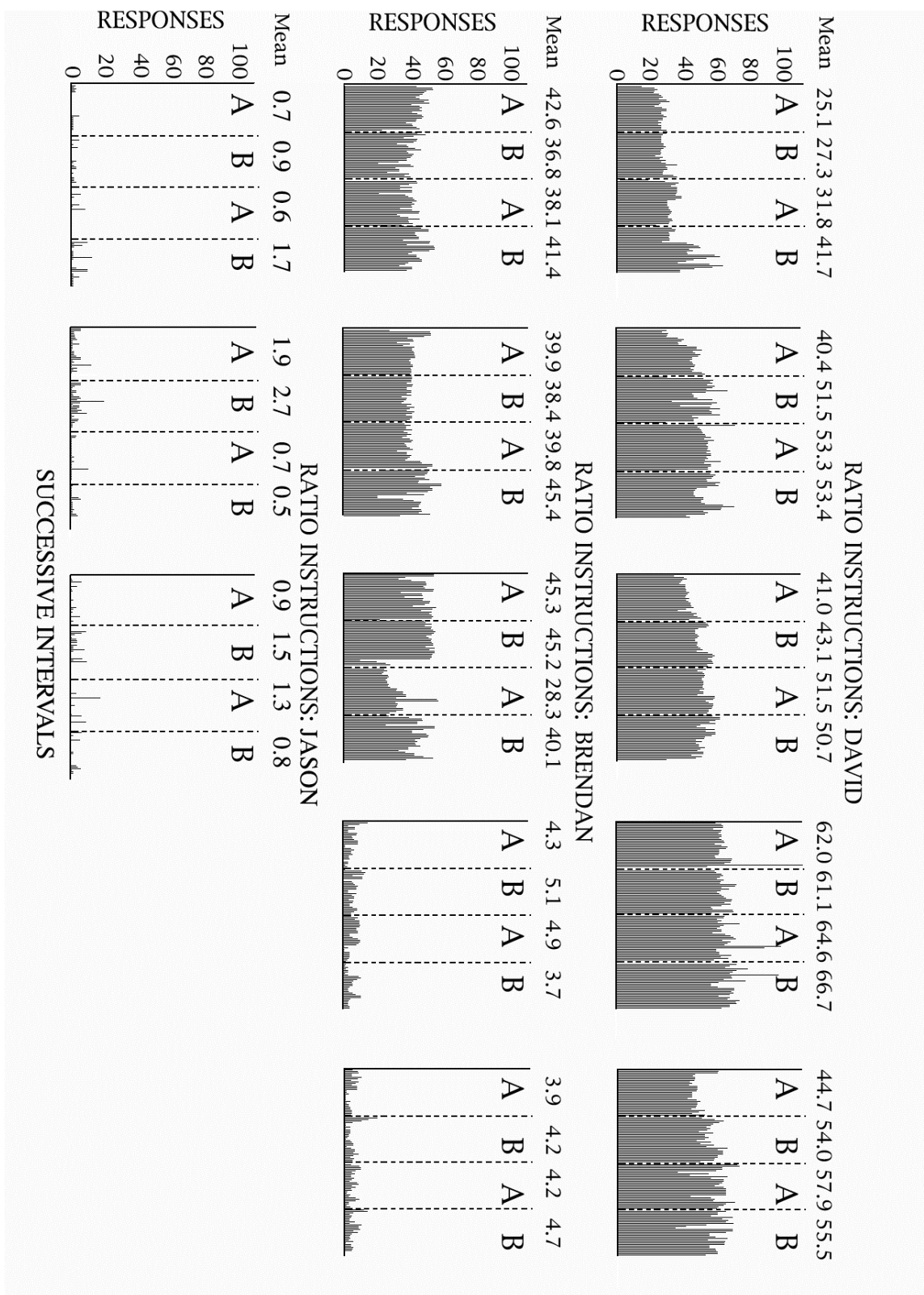


Figure 3. Number of responses during each FT component of Conditions A & B in each session for David, Brendan, and Jason.



Table 3

Ratio Instructions: ABAB

PRP durations in seconds for each subject across conditions and across sessions

	A		B		A		B	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
David	<b>0.55</b>	0.72	<b>0.39</b>	0.36	<b>0.24</b>	0.26	<b>0.14</b>	0.17
	<b>0.17</b>	0.20	<b>0.12</b>	0.13	<b>0.13</b>	0.15	<b>0.14</b>	0.15
	<b>0.17</b>	0.16	<b>0.12</b>	0.87	<b>0.12</b>	0.12	<b>0.12</b>	0.12
	<b>0.09</b>	0.10	<b>0.08</b>	0.09	<b>0.07</b>	0.07	<b>0.07</b>	0.07
	<b>0.14</b>	0.14	<b>0.09</b>	0.10	<b>0.09</b>	0.12	<b>0.09</b>	0.11
Brendan	<b>0.57</b>	1.07	<b>0.66</b>	1.03	<b>0.22</b>	0.74	<b>0.22</b>	0.59
	<b>0.17</b>	0.23	<b>0.25</b>	0.23	<b>0.14</b>	0.23	<b>0.13</b>	0.20
	<b>0.12</b>	0.21	<b>0.12</b>	0.50	<b>0.09</b>	0.16	<b>0.16</b>	0.19
	<b>3.02</b>	4.48	<b>1.92</b>	2.85	<b>2.20</b>	3.76	<b>0.63</b>	2.12
	<b>3.47</b>	4.27	<b>3.50</b>	4.00	<b>2.25</b>	3.93	<b>2.10</b>	3.00
Jason	<b>21.18</b>	10.72	<b>9.67</b>	9.82	<b>13.08</b>	10.74	<b>11.50</b>	10.67
	<b>9.87</b>	9.12	<b>9.73</b>	9.75	<b>11.73</b>	13.18	<b>11.47</b>	13.39
	<b>11.80</b>	15.50	<b>15.54</b>	14.62	<b>12.42</b>	14.65	<b>11.42</b>	12.42

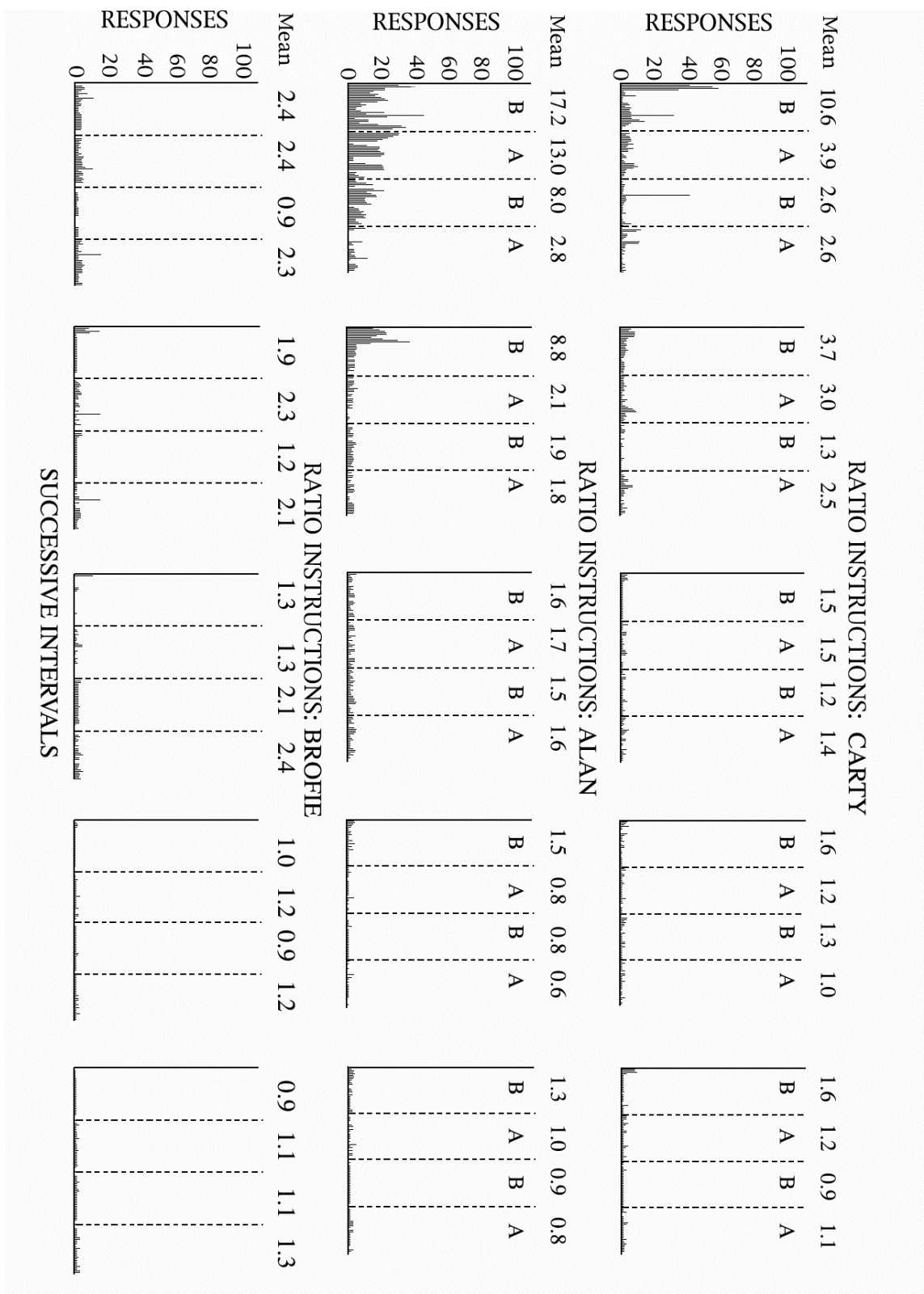


Figure 4. Number of responses during each FT component of Conditions A & B in each session for Carty, Alan, and Brofie.

Table 4

## Ratio Instructions: BABA

PRP durations in seconds for each subject across conditions and across sessions

	B		A		B		A	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
Carty	<b>4.71</b>	5.27	<b>3.59</b>	4.43	<b>1.42</b>	2.64	<b>6.77</b>	6.14
	<b>3.33</b>	3.79	<b>4.76</b>	4.90	<b>4.53</b>	4.71	<b>5.97</b>	5.76
	<b>1.92</b>	2.79	<b>6.68</b>	6.69	<b>5.52</b>	5.33	<b>8.53</b>	7.89
	<b>4.21</b>	4.79	<b>8.99</b>	7.40	<b>9.17</b>	7.86	<b>9.51</b>	9.07
	<b>9.04</b>	7.14	<b>9.47</b>	8.88	<b>9.40</b>	9.44	<b>9.30</b>	9.18
Alan	<b>1.26</b>	1.73	<b>0.67</b>	2.66	<b>1.16</b>	2.94	<b>8.96</b>	7.64
	<b>4.82</b>	5.17	<b>8.41</b>	8.21	<b>8.00</b>	8.06	<b>8.24</b>	8.23
	<b>8.97</b>	8.78	<b>8.91</b>	8.55	<b>9.27</b>	8.98	<b>9.42</b>	8.75
	<b>9.37</b>	8.90	<b>9.83</b>	9.59	<b>9.73</b>	9.87	<b>10.10</b>	9.85
	<b>9.34</b>	8.98	<b>9.71</b>	9.34	<b>9.36</b>	9.52	<b>9.92</b>	9.57
Brofie	<b>1.94</b>	4.49	<b>9.08</b>	8.47	<b>8.56</b>	5.64	<b>8.85</b>	7.35
	<b>0.55</b>	1.0	<b>8.83</b>	7.02	<b>0.62</b>	1.81	<b>8.94</b>	5.57
	<b>0.54</b>	1.08	<b>0.84</b>	4.79	<b>0.40</b>	0.74	<b>8.62</b>	5.97
	<b>0.36</b>	1.76	<b>0.50</b>	1.69	<b>0.56</b>	2.72	<b>0.44</b>	0.46
	<b>0.61</b>	1.21	<b>0.50</b>	1.40	<b>0.36</b>	0.73	<b>0.41</b>	0.42

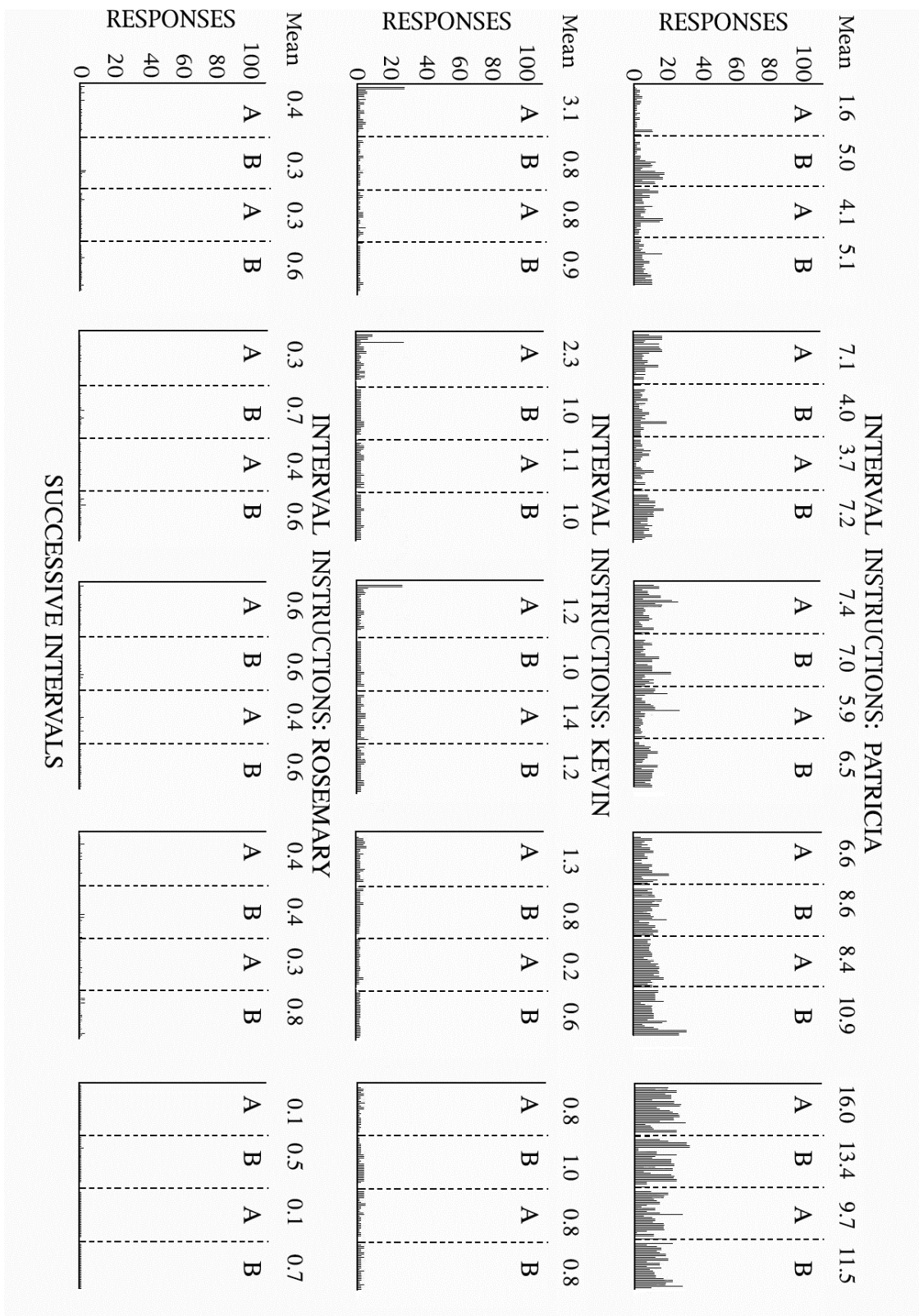


Figure 5. Number of responses during each FT component of Conditions A & B in each session for Patricia, Kevin, and Rosemary.

Table 5

## Interval Instructions: ABAB

PRP durations in seconds for each subject across conditions and across sessions

	A		B		A		B	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
Patricia	<b>7.21</b>	7.35	<b>7.22</b>	6.26	<b>8.82</b>	7.56	<b>7.60</b>	7.36
	<b>7.23</b>	6.06	<b>8.73</b>	8.25	<b>8.57</b>	8.47	<b>7.29</b>	6.98
	<b>7.48</b>	6.95	<b>7.39</b>	6.69	<b>8.42</b>	7.63	<b>7.40</b>	7.54
	<b>6.87</b>	6.72	<b>6.60</b>	6.93	<b>6.78</b>	7.18	<b>6.62</b>	6.24
	<b>2.17</b>	3.33	<b>1.77</b>	3.88	<b>6.63</b>	6.08	<b>6.29</b>	5.99
Kevin	<b>7.24</b>	6.59	<b>8.73</b>	8.35	<b>9.62</b>	9.61	<b>0.77</b>	3.45
	<b>8.32</b>	7.0	<b>0.92</b>	2.95	<b>6.79</b>	6.83	<b>1.09</b>	2.28
	<b>6.59</b>	6.66	<b>0.60</b>	1.92	<b>5.67</b>	5.22	<b>0.58</b>	2.08
	<b>5.63</b>	6.64	<b>5.28</b>	6.62	<b>15.67</b>	14.93	<b>8.53</b>	9.82
	<b>9.35</b>	9.81	<b>0.81</b>	6.65	<b>8.92</b>	9.76	<b>3.05</b>	7.46
Rosemary	<b>11.90</b>	10.64	<b>13.12</b>	12.46	<b>11.20</b>	11.16	<b>9.50</b>	9.67
	<b>10.93</b>	11.02	<b>8.48</b>	8.92	<b>11.09</b>	10.92	<b>10.53</b>	10.90
	<b>9.76</b>	9.87	<b>9.92</b>	9.55	<b>10.58</b>	10.56	<b>9.69</b>	10.05
	<b>10.63</b>	11.43	<b>10.94</b>	10.88	<b>11.56</b>	11.41	<b>9.39</b>	9.70
	<b>12.05</b>	11.88	<b>11.07</b>	10.90	<b>11.15</b>	11.11	<b>9.49</b>	9.89



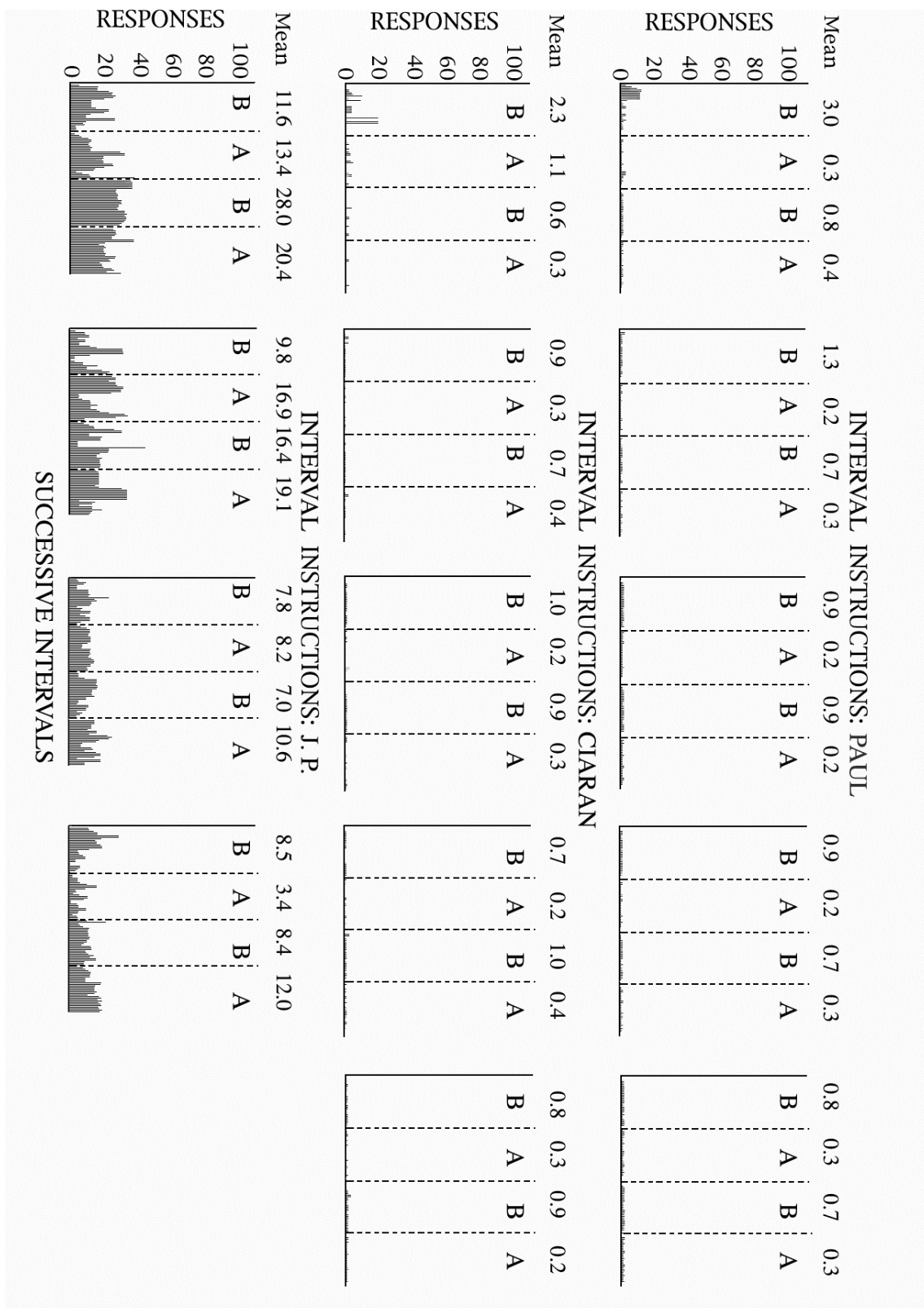


Figure 6. Number of responses during each FT component of Conditions A & B in each session for Paul, Ciaran, and J.P.

Table 6

Interval Instructions: BABA

PRP durations in seconds for each subject across conditions and across sessions

	B		A		B		A	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
Paul	<b>8.88</b>	7.49	<b>11.65</b>	11.04	<b>0.98</b>	3.47	<b>10.43</b>	8.43
	<b>1.21</b>	1.40	<b>10.96</b>	10.40	<b>1.70</b>	4.96	<b>10.54</b>	10.03
	<b>0.54</b>	2.26	<b>11.44</b>	9.73	<b>0.65</b>	1.61	<b>11.41</b>	9.65
	<b>0.46</b>	0.94	<b>11.32</b>	9.70	<b>9.46</b>	5.29	<b>10.04</b>	9.77
	<b>0.47</b>	0.23	<b>10.27</b>	7.53	<b>0.77</b>	3.82	<b>10.24</b>	7.12
Ciaran	<b>1.13</b>	2.28	<b>8.58</b>	8.75	<b>1.03</b>	10.68	<b>10.39</b>	11.60
	<b>1.44</b>	4.10	<b>10.56</b>	9.72	<b>1.83</b>	4.98	<b>10.82</b>	9.61
	<b>0.79</b>	1.49	<b>10.82</b>	10.28	<b>0.90</b>	2.54	<b>10.67</b>	9.33
	<b>1.44</b>	5.03	<b>11.43</b>	10.73	<b>0.76</b>	1.76	<b>10.60</b>	10.41
	<b>1.66</b>	5.10	<b>11.01</b>	11.23	<b>14.01</b>	2.45	<b>10.81</b>	10.67
J.P.	<b>0.43</b>	0.52	<b>0.39</b>	0.44	<b>0.26</b>	0.24	<b>0.40</b>	0.39
	<b>1.08</b>	2.25	<b>0.25</b>	0.51	<b>0.39</b>	0.70	<b>0.46</b>	0.77
	<b>0.97</b>	1.97	<b>0.77</b>	1.38	<b>0.57</b>	1.99	<b>0.41</b>	0.70
	<b>0.62</b>	2.51	<b>2.36</b>	4.59	<b>0.83</b>	1.25	<b>0.54</b>	0.60