Journal of Consulting and Clinical Psychology
1994, Vol. 62, No. 5, 942-948

0022-006X/94/$3.00

Disabuse of the Drug Metaphor: Psychotherapy Process–Outcome Correlations

William B. Stiles and David A. Shapiro

The drug metaphor suggests that if a process component (e.g., interpretation) is an active ingredient of a successful psychotherapy, then administering a relatively high level of it should yield a relatively positive outcome, and levels of the process component and the outcome should be correlated across clients. Measures of 5 theoretically relevant, reliably measured verbal process components were compared with the rate of change in 3 standard symptom intensity measures across the brief treatments of 39 (mainly depressed) psychotherapy clients. The expected significant process–outcome correlations were not found. These results are discussed as they illuminate some misleading assumptions that underlie many conventional studies of psychotherapy process and outcome.

How does psychotherapy work? How do the conversations between therapists and clients (psychotherapy process) reduce psychological suffering and promote productive, satisfying ways of living (psychotherapy outcome)?

The traditional approach to assessing process–outcome relationships derives from an implicit (or sometimes explicit) assumption, which we call the *drug metaphor* (Stiles & Shapiro, 1989), that the "ingredients" of verbal psychotherapy can be treated like the ingredients of pharmacological therapies in evaluation of their strength, integrity, and effectiveness (cf. Yeaton & Sechrest, 1981). The ingredients are process components—verbal and nonverbal productions by the client and interventions by the therapist. If a component is an active ingredient, then administering a high level of it is supposed to yield a positive outcome. If it does not, the process component is presumed to be inert.

Although reviewers agree that the ideal way to assess the efficacy of process components would be by controlled experiment, it would be prohibitively difficult and expensive to manipulate a single component of psychotherapy process (e.g., number of interpretations) while holding other factors constant across randomly assigned therapist–client dyads in a clinical trial. Instead, most studies have measured naturally occurring variation in process and outcome and assessed their intercorrelation (Orlinsky & Howard, 1986). The correlational corollary of the drug metaphor model suggests that if a process component is an active ingredient, then clients who receive relatively more of it should tend to improve more, so measures of this process should be positively correlated with measures of outcome across clients. We applied this logic to data from the first Sheffield Psychotherapy Project (Shapiro & Firth, 1987).

For process components, we focused on a few verbal categories measured by the verbal response mode (VRM) coding system (Stiles, 1986, 1992). Obviously, VRMs in general, and particularly the categories we selected, represent only a few facets of the multifaceted therapeutic process. However, there are reasons to believe that some VRMs are important in therapy, so it is reasonable to apply tests implied by the drug metaphor logic.

Method

Design, Participants, and Outcomes

Each of 39 clients underwent an intake assessment and was randomly assigned to 8 weekly sessions of either a psychodynamic–interpersonal treatment (PI, also called exploratory therapy; $n = 18$) or a cognitive–behavioral treatment (CB, also called prescriptive therapy; $n = 21$). Characterizations of these treatments have been published previously (Hobson, 1985; Shapiro & Firth, 1987; Stiles, Shapiro, & Firth-Cozens, 1988) and brief project treatment manuals are available (Firth & Shapiro, 1985; Shapiro & Firth, 1985). After a midtherapy assessment, clients received the other treatment (CB or PI) with the same therapist for eight further weekly sessions, followed by termination and follow-up assessments. The order of treatments was counterbalanced within therapists in this crossover design. All sessions were audiotape recorded. Half of each client's tapes were VRM coded. One additional client might have been recognized by the coders and was excluded.

The clients were 23 men and 16 women in professional and managerial occupations; their mean age was 40. Thirty were diagnosed depressed by the PSE-ID-CATEGO system (Present State Examination Index of Definition Category System; Wing, Cooper, & Sartorius, 1974); the rest had mainly anxiety disorders. The therapists were four clinical psychologists (one man, three women) who were also investigators on the project.

Standard outcome instruments were administered at intake, midtherapy, termination, and at 3-month and 2-year follow-ups. Measures included the Present State Examination (PSE; Wing et al., 1974), which is based on a structured interview, and self-report inventories, including

William B. Stiles, Department of Psychology, Miami University; David A. Shapiro, Medical Research Council/Economic and Social Research Council Social and Applied Psychology Unit, University of Sheffield, United Kingdom.

Portions of this article were reported in June 1989 at the Invitational Conference on Interaction Process and Analysis, Kill Devil Hills, North Carolina, and in September 1989 at the European Conference of the Society for Psychotherapy Research, Bern, Switzerland.

This research was supported in part by Grant MH-40503 from the National Institute of Mental Health.

Correspondence concerning this article should be addressed to William B. Stiles, Department of Psychology, Miami University, Oxford, Ohio 45056. Electronic mail may be sent to wbstiles-@miamiu.acs.muohio.edu.

the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) and a 90-item Symptom Checklist (SCL-90; Derogatis, Lipman, & Covi, 1973).

Clients made substantial gains during therapy and maintained these at 3-month and 2-year follow-ups (Shapiro & Firth, 1987; Shapiro & Firth-Cozens, 1990). The improvement was clinically as well as statistically significant; about half the clients scored more than two standard deviations below the intake mean on symptom measures at 3-month follow-up, the most rigorous of the criteria discussed by Jacobson, Follette, and Revensdorf (1984).

VRM Coding

For each client, eight complete sessions were coded according to the VRM taxonomy (Stiles, 1992)—four sessions from each eight-session treatment period, normally the first, third, fifth, and eighth. Thus, a total of 312 sessions were studied. Coders worked directly from audio recordings. Details of the VRM coding procedures and reliability are reported elsewhere (Stiles et al., 1988).

The VRM system classifies each utterance according to its grammatical form and its pragmatic intent. As used here, the VRM taxonomy included 64 possible categories of client utterances (8 forms × 8 intents), and 80 possible categories of therapist utterances (8 forms × 10 intents; two of the standard intent categories were subdivided to study theoretical issues). These categories can be combined to yield many indexes to characterize the verbal interaction (Stiles, 1986, 1992).

For this report, we selected five VRM indexes that were of particular theoretical interest—four therapist indexes and one client index. Each of these was collapsed across VRM forms to yield an index of VRM intent. (VRM intent refers to the pragmatic meaning of an utterance, and coding it does not require the coder to read the speaker's mind; Stiles, 1987). The therapist indexes were Question, General Advisement, Interpretation, and Exploratory Reflection intents. The client index was Disclosure intent. (We capitalize category names to distinguish them from colloquial uses of the terms.)

Therapist Question intents are information-seeking utterances. They are important in CB in probing for detailed information about the nature and circumstances of the presenting problems—exactly when and how the difficulties are manifested. They are discouraged in PI because they may limit the client's freedom of action in responding, by imposing the therapist's frame of reference, and risk raising client anxiety by demanding a specific response (Hobson, 1985). The Question intent index does not include Question forms used in service of other modes, such as Interpretation or Advisement.

Therapist General Advisement intents are directive utterances aiming to guide the client's behavior outside the session. In the Sheffield study, they were distinguished from another subdivision of directives, called Process Advisements, which aim to guide the client's in-session behavior. General Advisements are theoretically important in CB, in which they are used to give homework assignments and to suggest strategies for overcoming problems. They are proscribed in PI, where the therapist's role is to aid clients' understanding of their relationship problems rather than to recommend solutions.

Therapist Interpretation intents put the client's experience into a frame of reference provided by the therapist—that is, they explain the client to himself or herself, label aspects of the client's thinking or acting, or reframe the client's experience. Interpretation has a place in both PI and CB (Elliott et al., in press), but it is particularly important in PI, in which reaching revised ways of understanding one's interpersonal relationships is seen as central. Interpretation in CB includes labeling client behaviors according to CB principles. In empirical studies of perceptions of therapy session dialogue, clients, therapists, and observers have tended to rate Interpretations as more helpful than other therapist interventions (Elliott, Barker, Caskey, & Pistrang, 1982; Hill et al., 1988).

Therapist Exploratory Reflection intents are "deep" Reflections, distinguished from Interpretations by their use of the client's, rather than the therapist's, frame of reference (Stiles, 1992). Although its content may go beyond what the client has literally said, an Exploratory Reflection is tentative, and the client is presumed to be the final arbiter of its accuracy. Exploratory Reflections were distinguished from Simple Reflections, which repeat the meanings (although not necessarily the words) already explicitly expressed by the client (Barkham & Shapiro, 1986). They are of great importance in PI, in which they encourage self-examination and articulation of buried feelings and attitudes. Although not explicitly proscribed, Exploratory Reflections have little place in CB, in which digging for deeper meanings is irrelevant.

Client Disclosure intents are utterances that reveal the client's private experience (e.g., thoughts, feelings, wishes, perceptions, and intentions), which could not be known directly by an external observer. Theoretically, client self-disclosure is a central process in all therapy. Research on VRM Disclosure suggests that it represents a substantial proportion of client talk regardless of therapy type, making it part of a common core of psychotherapy process (Stiles & Sultan, 1979; Stiles et al., 1988). The percentage of Disclosure in a client's speech is highly correlated with expert ratings of "good process" as measured by the Experiencing scale (Stiles, McDaniel, & McGaughey, 1979) and the Therapist and Patient Exploration indexes of the Vanderbilt Psychotherapy Process Scale (McDaniel, Stiles, & McGaughey, 1981).

Results

VRM Differences Between Treatments

Consistent with theoretical expectations, there were large differences between treatments in therapists' VRM use (mean percentage of therapist or client utterances in each category in each session; see Table 1; see also Stiles et al., 1988): more Question and General Advisement in CB versus more Interpretation and Exploratory Reflection in PI. Client Disclosure was significantly higher in PI than in CB sessions but was the most common client VRM intent in both treatments. Intercoder reliability was measured as the intraclass correlation coefficient designated ICC (1, 1) by Shrout and Fleiss (1979), in which coder differences are treated as random effects, and was at least adequate for all five indexes (Table 1).

Decay Curves as a Measure of Change

Raw change scores are controversial as an index of therapeutic improvement (Cronbach & Furby, 1970; Rogosa & Willett, 1985; Willett, 1988). An alternative analytic approach is that of "growth curves" (Francis, Fletcher, Stuebing, Davidson, & Thompson, 1991; Rogosa & Willett, 1985; Willett, Ayoub, & Robinson, 1991), which are more appropriately called *decay curves* when applied to change in symptom intensity across psychotherapeutic treatment. According to this approach, change is considered as a continuous process underlying performance on multiple assessments. A separate decay curve is fitted to each client's series of scores on an assessment measure, and some descriptive parameter (typically the curve's slope) is used as a quantitative index of that client's change on that measure.

For each client, we fitted regression lines for each of the three assessment instruments (PSE, BDI, SCL-90) across Assessments 1, 2, and 3 (intake, midtherapy, and termination), and used the slopes of these decay curves as indexes of change in therapy. Mean slopes were −3.9 (SD = 2.8) for the PSE, −3.7

Table 1
Means of Verbal Response Mode Percentages in Psychodynamic–Interpersonal and Cognitive–Behavioral Psychotherapies and Correlations With Rate of Change on Assessment Measures

Statistic	Therapist VRM intent				Client Disclosure intent
	Question	General Advisement	Interpretation	Exploratory Reflection	
Reliability[a]	.90	.91	.84	.75	.85
Mean percentage of each speaker's utterances					
PI	5.7	0.8	20.3	8.3	42.9
CB	10.1	7.0	14.3	1.9	36.6
Treatment main effect					
$F(1, 37)$[b]	36.63**	282.73**	47.63**	144.89**	31.31**
Correlation of VRM mean percentage in all coded sessions with rate of change from intake to termination					
PSE	.06	.01	−.11	.11	.02
BDI	.06	−.05	.08	.18	.23
SCL-90	.21	.09	−.07	.05	.07
Correlation of VRM mean percentage in PI sessions with rate of change across PI period					
PSE	−.37*	.25	.06	−.03	−.15
BDI	−.25	.02	.08	.14	.04
SCL-90	−.18	.02	.00	.06	−.03
Correlation of VRM mean percentage in CB sessions with rate of change across CB period					
PSE	.11	.21	−.16	.08	−.04
BDI	.07	.06	.05	.10	−.08
SCL-90	.17	.07	−.02	.06	−.12

Note. $N = 39$ clients for means. $N = 37$ to 39 clients for correlations because of missing data on some measures. Table entries for the separate treatments are correlations of residuals, corrected for order-of-treatment effects. Negative correlations indicate that higher verbal response mode (VRM) percentages were associated with declining scores on assessment measures (i.e., with symptomatic improvement). PI = psychodynamic–interpersonal treatment; CB = cognitive–behavioral treatment; PSE = Present State Examination (Wing et al., 1974); BDI = Beck Depression Inventory (Beck et al., 1961); SCL-90 = Symptom Checklist–90 (Derogatis et al., 1973).
[a] Reliability = intercoder reliability of a single coder's percentage of each code in a session, measured as the intraclass correlation coefficient designated ICC(1,1) by Shrout and Fleiss (1979) calculated across 84 sessions coded independently by two coders. [b] The F values represent univariate tests of the within-subjects treatment factor in two-factor analyses of variance; the between-subjects order-of-treatments factor and the Treatment × Order interaction had few significant effects and are not shown.
* $p < .05$. ** $p < .001$.

($SD = 2.5$) for the BDI, and −24.7 ($SD = 17.6$) for the SCL-90. The slopes thus represented each client's rate of change on each instrument, taking into account all three assessments, rather than only the beginning and end points.

Process–Outcome Correlations

The five VRM indexes did not predict clients' improvement in therapy. Correlations of the VRM indexes (mean percentages across all coded sessions) with outcomes, measured as slope of the decay curve across the three assessments on the PSE, BDI, and SCL-90, were all nonsignificant (Table 1). Correlations of these VRM indexes with other change indexes, including decay curve slopes across all four assessments (follow-up as well as intake, midtherapy, and termination), arithmetic change from intake to termination or intake to follow-up, and residual gains (Cronbach & Furby, 1970) from intake to termination or from intake to follow-up, were also all nonsignificant (data not shown).

Within-Treatment Process–Outcome Correlations

Because each therapist VRM variable was theoretically more important in one treatment than in the other, we also assessed the correlation of VRM use within each treatment (PI or CB) with change across that period (i.e., intake to midtherapy or midtherapy to termination). We calculated decay curves (slopes of regression lines) for each client on each assessment measure across each of the treatments separately. Because there were only two points for each slope, these were equivalent to arithmetic change scores. Because change tended to be greater in the first period than in the second period regardless of which treatment was first in the crossover design (Shapiro & Firth, 1987), we used residuals with respect to order of treatments; that is, rather than correlating the VRM percentages and slopes themselves, we used each score's deviation from the mean for clients in that treatment order. As Table 1 shows, the correlations of the (residualized) VRM mean percentages with (residualized) slopes within each treatment were also essentially negligible.

Table 2
Correlations of Rate of Change in Verbal Response Mode Percentages With Rate of Change on Assessment Measures Across Psychodynamic–Interpersonal or Cognitive–Behavioral Psychotherapy Sessions

Statistic	Therapist VRM intent: Question	General Advisement	Interpretation	Exploratory Reflection	Client Disclosure intent
Correlation for PI period					
PSE	.31	−.01	.13	.09	−.11
BDI	.00	−.04	.24	−.20	.02
SCL-90	.20	−.08	.24	−.03	.12
Correlation for CB period					
PSE	.18	.08	.18	−.07	.25
BDI	.14	−.13	.11	.17	.37*
SCL-90	.22	−.04	.15	.32	.29

Note. N = 37 to 39 clients for correlations because of missing data on some measures. Table entries are correlations of residuals, corrected for order-of-treatment effects. Negative correlations indicate that increasing VRM percentages were associated with declining scores on assessment measures (i.e., with symptomatic improvement). VRM = verbal response mode; PI = psychodynamic–interpersonal treatment; PSE = Present State Examination (Wing et al., 1974); BDI = Beck Depression Inventory (Beck et al., 1961); SCL-90 = Symptom Checklist–90 (Derogatis et al., 1973); CB = cognitive–behavioral treatment.

One was nominally significant at the .05 level, but correction for multiple-comparison error rates would void this as being well within chance expectations.

Rate of Change in Process Versus Rate of Change in Symptom Intensity

Finally, we asked the question, were the growth or decay curves of the VRM percentages related to the decay curves of the assessment measures? To address this, for each client, we fitted regression lines for each VRM measure across the coded sessions of each treatment (four sessions for each client, each numbered between 1 and 8), and used the slope of the regression line as an index of change in VRM use within each treatment. As before, we next adjusted scores for order-of-treatment effects. We then correlated the (residualized) VRM slopes with the (residualized) assessment measure slopes across each treatment separately. (The large VRM differences between treatments, shown in Table 1, would render comparisons across both treatments uninterpretable.) A correlation between a VRM growth or decay curve and an assessment measure decay curve could represent participants' within-session verbal adaptation to clients' gradual improvement, or it could represent efficacy of a specific dynamic pattern of VRM use (incrementing or decrementing across sessions). These possibilities are somewhat different from the process–outcome relationship suggested by the drug metaphor but would nevertheless be of interest. However, as Table 2 shows, there was only one nominally significant correlation, a result which would not survive any Bonferroni or similar correction for multiple significance tests.

Discussion

The negligible process–outcome correlations do not appear to reflect a failure of measurement. On the process side, intercoder reliability was good, and the VRM indexes represented prominent, theoretically important components of the interaction that clearly discriminated between treatments in theoretically expected ways (Table 1; see also Stiles et al., 1988). On the outcome side, the PSE, BDI, and SCL-90 are widely used and respected measures that showed large, clinically and statistically significant changes in the Sheffield Project data (Shapiro & Firth, 1987). From evidence similar to this, reviewers often conclude that such verbal ingredients are ineffective (e.g., Beutler, Crago, & Arizmendi, 1986; Orlinsky & Howard, 1986).

Was this a fair test? The model represented by the correlations in Table 1 seems to be a straightforward quantitative representation of the drug metaphor conception, in which VRM percentages (process ingredients) were directly and independently compared with decay of symptom intensity across successive assessments (treatment outcomes). The essentially identical null correlations of the VRM variables with other indexes of improvement—difference scores and residual change scores—suggest that the lack of relationship does not reflect merely an unfortunate choice of an outcome statistic.

Furthermore, our results are consistent with the generally disappointing, inconsistent yield of psychotherapy process–outcome comparisons (Orlinsky & Howard, 1986; Shapiro et al., 1994). Although some studies report significant relationships, these are seldom reliably replicated. Typically, the best studies have found the fewest and weakest relationships. For example, Moras, Elkin, Imber, Watkins, and Sotsky (1989) developed a set of measures specifically tuned to the theory of Interpersonal therapy in the National Institute of Mental Health's collaborative project on the therapy of depression (Elkin et al., 1989). They measured both delivery of theoretically crucial components and uptake by the clients. After painstaking sampling and measurement (using a contrasting groups design, in which the processes in the therapies of good- and poor-outcome clients were compared), they too found negligible statistical relationships between process and outcome.

It would be possible, of course, to construct and test many alternative models of relationships among process and outcome

measures taken in the Sheffield Project, involving, for example, differently shaped decay curves or more complex combinations of process variables. The correlations shown in Table 2, which compared the rate of change in VRM process variables with the rate of change in assessment measures, represent one step in such a statistical exploration. We reported that particular analysis because it was recommended in a review of an earlier version of this article. However, searching through multiple analytic strategies presents the same threats to statistical conclusion validity as does searching through multiple significance tests. Our purpose in this article was not to seek significant relationships among Sheffield Project measures but to confront the process–outcome correlation problem directly.

So, do the negligible process–outcome correlations (Table 1) show that VRM-measured process components are inert ingredients of psychotherapy? We think not. We suggest instead that the underlying drug metaphor is misleading (Stiles & Shapiro, 1989). In particular, we believe that the logic underlying this interpretation of the correlations is flawed and that process–outcome correlations cannot be trusted to reflect a process component's contribution to outcome (Stiles, 1988). We review our reasoning here, although, of course, our data alone do not prove that it is correct.

The process–outcome correlation logic overlooks therapist (and client) responsiveness to varying client requirements for process components. Obviously, there is variation across clients and across time in clients' requirements for particular types of verbal intervention. In suggesting that the strength (percentage, frequency, intensity) of active process ingredients should predict which clients improve, the logic makes the absurd assumption that process components are delivered randomly with respect to client requirements. Such a conversation would be incomprehensible.

Human interaction, including psychotherapy, is systematically appropriately responsive (e.g., Elliott, 1984; Elliott et al., 1990; Goodwin, 1981; Grice, 1975; Kent, Davis, & Shapiro, 1978; Labov & Fanshel, 1977). People answer each other's questions, stay on related topics, and take turns speaking using an elaborate system of signals. On top of normal conversational competence (and personal caring), much of therapy training and supervision is directed toward vigilance in sessions and adapting the type, depth, timing, and phrasing of interventions to the needs of particular clients at particular moments in treatment.

Any appropriate responsiveness tends to defeat the process–outcome correlation. In the limit, if responsiveness were perfect, then the expected correlation would be zero. Process component strength would vary in response to varying client requirements, and outcomes would vary for reasons other than a deficiency or excess of that component, but the process component and the outcome would not covary (at least not for reasons of the process component's efficacy) because each client would be receiving exactly enough. For example, a perfectly responsive therapist's questions in CB would vary with the therapist's need for information (e.g., for treatment planning) and with the quality of the client's answers. Improvement in BDI scores would vary with clients' life circumstances and capacity for change. But questions would be uncorrelated with BDI improvement, *even if asking questions is a crucial ingredient in the process of* CB. Asking additional questions would not improve any client's outcome.

Realistically, when responsiveness is substantial but imperfect, the correlation of crucial process components with outcomes may be misleadingly positive, null, or even negative. A hypothetical example can illuminate the problem: Suppose that, as the theory suggests, interpretation is an active ingredient in psychodynamic therapies. For many reasons, clients will vary in their requirements for interpretations. Some clients may be slow to understand or resistant to accepting important interpretations and may need to have them recast and repeated many times, whereas other clients may grasp and accept their therapists' interpretations quickly and so need fewer of them. Responsive therapists would adapt to these differing requirements, albeit imperfectly. Nevertheless, despite the therapists' efforts, resistant or recalcitrant clients (who received more interpretations) may be likely to have worse outcomes than quicker, more accepting clients (who received fewer). In this case, the incidence of interpretations would be negatively correlated with outcome across clients, although interpretation is an active ingredient in the treatment! This seemingly paradoxical result has been observed in some studies (e.g., Sloane, Staples, Cristol, Yorkston, & Whipple, 1975).

Even the interpretation example understates the problem by focusing on therapist responsiveness to stable client characteristics. People's conversational requirements change moment to moment, and interpersonal interaction is responsive on a millisecond time scale (Goodwin, 1981). For example, therapists often alter phrasing or amend or undo interventions based on client reactions that occur during the intervention (Elliott, 1984; Elliott et al., in press). Consequently, correlational tests of process–outcome relationships would be misleading even on an intervention-level time scale.

Arguments parallel to the foregoing question and interpretation examples can be constructed for virtually any process component (Stiles, 1988). The problem lies not in the choice or precision of the process measures (a perfectly measured essential component may have a null or negative correlation with outcome) but in the assumption of the linear statistical model that process components are administered randomly or independently of client requirements. This assumption is clearly violated in any psychotherapy.

To put this more generally, responsiveness implies that outcome feeds back to influence process. Therapists intervene because they think it will be beneficial, and they adjust their interventions moment by moment in response to the effects of previous interventions, so process components are contingent on (anticipated) outcomes or related cues. (Note that responsiveness is construed as a process of acting on feedback, not as a therapist trait. It need not imply high activity or doing what pleases the client in the short term.) Feedback loops are implicit in such theoretical concepts as empathy, timing, and countertransference.

Feedback can render relationships between variables chaotic even in very simple systems; that is, relationships are likely to be unpredictable because of sensitive dependence on initial conditions (Gleick, 1987). Minor fluctuations in phrasing can initiate chains of reactions that may lead in unexpected directions and have large effects, for better or worse.

Insofar as feedback is characteristic of psychotherapy (indeed, of any human interaction), the process–outcome system is inherently nonlinear, and it is unlikely that correlations or related statistics, which are based on linear models, will adequately assess relationships between process and outcome. Note that the nonlinearity that leads to chaos (in the technical sense; Gleick, 1987) concerns the relationship between process and outcome, not with whether the growth or decay of process or of outcome itself is modeled as a straight line.

Other applications of linear models, such as contingencies within sequences of utterances (Russell & Trull, 1986) or relationships of utterance-level measures to session-level measures (e.g., Hill et al., 1988), may contribute to an understanding of psychotherapy process. Even here, however, responsiveness imposes limits. For example, because of feedback and sensitive dependence on initial conditions, utterance sequences cannot be predicted more than a few turns in advance, despite the common experience that some verbal interventions have powerful long-lasting effects and, hence, would loom large retrospectively in a qualitative or narrative analysis.

We are aware that the terms *process* and *outcome* can be used with varying scope. For example, each utterance can be described as an outcome of the preceding utterance; therapist and client postsession evaluations of their encounter can be called "micro-outcome" (Orlinsky & Howard, 1986) or "session outcome" (e.g., Hill et al., 1988). Such relabeling may have heuristic benefits, but it also may lead to confusion about what is really being studied. We believe it is important to confront the possibility that the conventionally understood drug-metaphor question of how process components (measured observationally within sessions) are related to outcome (measured as improvement across treatment) may be unanswerable within a conventional linear framework.

Paradoxically, finding flaws in the process-outcome correlation logic may restore confidence—or, at least, hope—in the efficacy of process components such as VRMs. One cannot legitimately infer from their negligible correlations with outcome that they are inert ingredients. Results such as those shown in Table 1 are entirely consistent with the theoretical propositions that therapist Questions, General Advisements, Interpretations, and Exploratory Reflections and client Disclosures, used responsively, are essential and potent components of PI and CB psychotherapy. The negligible correlations do not demonstrate efficacy, of course, but neither do they weigh against such indirect evidence as the theoretically coherent differential use of these VRMs (Table 1, upper section) in treatments that were, overall, highly efficacious (Shapiro & Firth, 1987). Methodological creativity will be needed to assess empirically the theoretical links between such process components and outcomes, but if the decades of discouraging failures to find consistent linear relationships can be disregarded, the effort again appears warranted.

References

Barkham, M., & Shapiro, D. A. (1986). Counselor verbal response modes and experienced empathy. *Journal of Counseling Psychology, 33,* 3–20.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4,* 561–571.

Beutler, L. E., Crago, M., & Arizmendi, T. G. (1986). Therapist variables in psychotherapy process and outcome. In S. L. Garfield & A. E. Bergin (Eds.), *Handbook of psychotherapy and behavior change* (3rd ed., pp. 257–310). New York: Wiley.

Cronbach, L. J., & Furby, L. (1970). How should we measure 'change'—or should we? *Psychological Bulletin, 74,* 68–80.

Derogatis, L. R., Lipman, R. S., & Covi, L. (1973). SCL-90: An outpatient psychiatric rating scale—preliminary report. *Psychopharmacology Bulletin, 9,* 13–27.

Elkin, I., Shea, M. T., Watkins, J. T., Imber, S. D., Sotsky, S. M., Collins, J. F., Glass, D. R., Pilkonis, P. A., Leber, W. R., Docherty, J. P., Fiester, S. J., & Parloff, M. B. (1989). National Institute of Mental Health treatment of depression collaborative research program: General effectiveness of treatments. *Archives of General Psychiatry, 46,* 971–982.

Elliott, R. (1984). A discovery-oriented approach to significant change events in psychotherapy: Interpersonal process recall and comprehensive process analysis. In L. N. Rice & L. S. Greenberg (Eds.), *Patterns of change* (pp. 249–286). New York: Guilford.

Elliott, R., Barker, C. B., Caskey, N., & Pistrang, N. (1982). Differential helpfulness of counselor verbal response modes. *Journal of Counseling Psychology, 29,* 354–361.

Elliott, R., Shapiro, D. A., Firth-Cozens, J., Stiles, W. B., Hardy, G. E., Llewelyn, S. P., & Margison, F. R. (in press). Comprehensive process analysis of insight events in cognitive-behavioral and interpersonal-dynamic psychotherapies. *Journal of Counseling Psychology.*

Firth, J. A., & Shapiro, D. A. (1985). *Prescriptive therapy manual for the Sheffield Psychotherapy Project* (Memo No. 734). (Available from Medical Research Council/Economic and Social Research Council Social and Applied Psychology Unit, University of Sheffield, Sheffield S10 2TN United Kingdom.)

Francis, D. J., Fletcher, J. M., Stuebing, K. K., Davidson, K. C., & Thompson, N. M. (1991). Analysis of change: Modeling individual growth. *Journal of Consulting and Clinical Psychology, 59,* 27–37.

Gleick, J. (1987). *Chaos: Making a new science.* New York: Penguin.

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers.* San Diego, CA: Academic Press.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics, Vol. 3: Speech acts.* San Diego, CA: Academic Press.

Hill, C. E., Helms, J. E., Tichenor, V., Spiegel, S. B., O'Grady, K. E., & Perry, E. (1988). The effects of therapist response modes in brief psychotherapy. *Journal of Counseling Psychology, 35,* 222–233.

Hobson, R. F. (1985). *Forms of feeling.* London: Tavistock.

Jacobson, N. S., Follette, W. C., & Revensdorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and clinical significance. *Behavior Therapy, 15,* 336–352.

Kent, G. S., Davis, J. D., & Shapiro, D. A. (1978). Resources required in the construction and reconstruction of conversation. *Journal of Personality and Social Psychology, 36,* 13–22.

Labov, W., & Fanshel, D. (1977). *Therapeutic discourse.* San Diego, CA: Academic Press.

McDaniel, S. H., Stiles, W. B., & McGaughey, K. J. (1981). Correlations of male college students' verbal response mode use in psychotherapy with measures of psychological disturbance and psychotherapy outcome. *Journal of Consulting and Clinical Psychology, 49,* 571–582.

Moras, K., Elkin, I., Imber, S., Watkins, J., & Sotsky, S. (1989, May). *The therapeutic action of interpersonal psychotherapy.* Paper presented at the meeting of the American Psychiatric Association.

Orlinsky, D. E., & Howard, K. I. (1986). Process and outcome in psychotherapy. In S. L. Garfield & A. E. Bergin (Eds.), *Handbook of psy-*

chotherapy and behavior change (3rd ed., pp. 311–381). New York: Wiley.

Rogosa, D. R., & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika, 50,* 203–228.

Russell, R. L., & Trull, T. J. (1986). Sequential analyses of language variables in psychotherapy process research. *Journal of Consulting and Clinical Psychology, 54,* 16–21.

Shapiro, D. A., & Firth, J. A. (1985). *Exploratory therapy manual for the Sheffield Psychotherapy Project* (Memo No. 733). (Available from Medical Research Council/Economic and Social Research Council Social and Applied Psychology Unit, University of Sheffield, Sheffield S10 2TN United Kingdom.)

Shapiro, D. A., & Firth, J. (1987). Prescriptive vs. exploratory psychotherapy: Outcomes of the Sheffield Psychotherapy Project. *British Journal of Psychiatry, 151,* 790–799.

Shapiro, D. A., & Firth-Cozens, J. A. (1990). Two year follow-up of the Sheffield Psychotherapy Project. *British Journal of Psychiatry, 157,* 389–391.

Shapiro, D. A., Harper, H., Startup, M. J., Reynolds, S., Bird, D., & Suokas, A. (1994). The high water mark of the drug metaphor. In R. L. Russell (Ed.), *Reassessing psychotherapy research* (pp. 1–35). New York: Guilford Press.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420–428.

Sloane, R. G., Staples, F. R., Cristol, A. H., Yorkston, N. J., & Whipple, K. (1975). *Psychotherapy versus behavior therapy.* Cambridge, MA: Harvard University Press.

Stiles, W. B. (1986). Development of taxonomy of verbal response modes. In L. S. Greenberg & W. M. Pinsof (Eds.), *The psychotherapeutic process: A research handbook* (pp. 161–199). New York: Guilford Press.

Stiles, W. B. (1987). Some intentions are observable. *Journal of Counseling Psychology, 34,* 236–239.

Stiles, W. B. (1988). Psychotherapy process-outcome correlations may be misleading. *Psychotherapy, 25,* 27–35.

Stiles, W. B. (1992). *Describing talk: A taxonomy of verbal response modes.* Newbury Park, CA: Sage.

Stiles, W. B., McDaniel, S. H., & McGaughey, K. J. (1979). Verbal response mode correlates of experiencing. *Journal of Consulting and Clinical Psychology, 47,* 795–797.

Stiles, W. B., & Shapiro, D. A. (1989). Abuse of the drug metaphor in psychotherapy process-outcome research. *Clinical Psychology Review, 9,* 521–543.

Stiles, W. B., Shapiro, D. A., & Firth-Cozens, J. A. (1988). Verbal response mode use in contrasting psychotherapies: A within-subjects comparison. *Journal of Consulting and Clinical Psychology, 56,* 727–733.

Stiles, W. B., & Sultan, F. E. (1979). Verbal response mode use by clients in psychotherapy. *Journal of Consulting and Clinical Psychology, 47,* 611–613.

Willett, J. B. (1988). Questions and answers in the measurement of change. In E. Z. Rothkopf (Ed.), *Review of research in education* (Vol. 15, pp. 345–422). Washington, DC: American Educational Research Association.

Willett, J. B., Ayoub, C. C., & Robinson, D. (1991). Using growth modeling to examine systematic differences in growth: An example of change in the functioning of families at risk of maladaptive parenting, child abuse, or neglect. *Journal of Consulting and Clinical Psychology, 59,* 38–47.

Wing, J. K., Cooper, J. E., & Sartorius, N. (1974). *The measurement and classification of psychiatric symptoms.* Cambridge, England: Cambridge University Press.

Yeaton, W. H., & Sechrest, L. (1981). Critical dimensions in the choice and maintenance of successful treatments: Strength, integrity, and effectiveness. *Journal of Consulting and Clinical Psychology, 49,* 156–167.

Received May 8, 1991
Revision received February 18, 1992
Accepted October 12, 1993 ■