

Data-Based Management for Tracking Outcome in Private Practice

Michael J. Lambert, Brigham Young University
G. S. (Jeb) Brown, Human Affairs International

Clement has proposed several categories of data to be included in data-based management of clinical practice for the purpose of improving quality of patient care. His model for routine evaluation includes a proposed system for evaluating treatment outcome. An alternative system for evaluating outcome is suggested by the authors. Its presumed value is greater efficiency in clinician (and patient) time involvement, more likelihood of contributing to the empirical understanding of therapeutic effects, and greater ease of comparison with the data produced from controlled research and from other private practitioners. Sample data are presented illustrating the value of using cutoff scores and social validation methodology to improve the quality of psychotherapeutic services and increase our understanding of the impacts of treatment.

Key words: psychotherapy outcome, private practice, assessment, clinical significance. [*Clin Psychol Sci Prac* 3:172-178, 1996]

Clement (1996) is to be congratulated for addressing the important issue of quality assessment and assurance by the clinician in routine clinical practice. The system of evaluation he proposes includes outcome assessment as its central focus with the goal of providing data that can be summed across patient diagnostic categories, treatment methods, and clinicians. He presents instructions for calculating a treatment effect size for each patient. He suggests the development of a database management system by individual providers and provider groups. This commentary supports the recommendations of Clement, offers some criticisms of his proposal, and makes suggestions for improved procedures.

Address correspondence to Michael J. Lambert, Department of Psychology, 272 TLRB, Brigham Young University, Provo, UT 84602.

CRITICAL EVALUATION OF CLEMENT'S OUTCOME MEASUREMENT RECOMMENDATIONS

The outcome measurement method proposed by Clement is based on formulating individualized goals based on Goal Attainment scaling (Kiresuk and Sherman, 1968) and Global Assessment of Functioning (American Psychiatric Association, 1987). Evaluating patient outcome on the basis of individualized goal attainment has the advantage of being particularly sensitive to the concerns of the individual patient while reflecting the results of therapist and patient collaborative efforts in therapy. In fact, individualized goal evaluation was recommended frequently in the late 1960s, with a resurgence of interest in the 1990s related to the advocacy of qualitative research methods by psychotherapy researchers (Lambert and Hill, 1994). Numerous procedures for individualizing treatment goals have been promoted over the years, with the Target Complaints method (Battle et al., 1966) and Goal Attainment scaling being the most widely used in treatment research. The Target Complaints method was a recommended outcome measure in the Core Outcome Battery as suggested by the National Institute of Mental Health core battery conference (Waskow and Parloff, 1975). Despite the advocacy of such methods by some, individualized goals have many limitations, some of which were alluded to by Clement. However, these methods have numerous shortcomings beyond those mentioned by Clement:

1. Therapist treatment effects are difficult to separate from therapist goal-setting bias.
2. Goals judged either too easy to obtain or too difficult to obtain are often included for analysis.
3. Change is inevitably confounded with expectations for improvement so that effective treatment may be more a function of goal-setting standards than treatment itself.
4. Concurrent validity of goal attainment yields relatively low correlations with other outcome measures.
5. There have been a large number of modifications to goal attainment (the Clement proposal being just one recent variation), so it is hard to compare goal attainment methods from study to study (let alone practitioner to practitioner).
6. Goal setting is a time-intensive procedure, especially in the initial stages of clinical applications.
7. Individualization of goals may preclude the use of

case-mix adjustment statistics, a point we shall return to later.

Given these and related problems (see also the critique of Goal Attainment scaling by Mintz and Kiesler, 1982), the status of individualized goal-rating procedures remains tenuous. As Lambert and Hill (1994) have observed, "Effective individualization of goals for the purpose of assessing patient change remains an ideal rather than a reality. The intention to individualize goals is very appealing but the gap between intention and effective application appears to be rather large" (p. 79).¹

Despite their limitations, we do not favor elimination of the procedures for individualizing goals recommended by Clement. Rather, it seems desirable to supplement individualized goals with standardized clinical rating scales. Both individualized goals and standard rating scales should be used to evaluate treatment outcomes in clinical practice. Similar recommendations have been advanced by others who advocate integration of independent practice and clinical research (e.g., Marten and Heimberg, 1995).

USING STANDARDIZED SCALES AND CUTOFF SCORES FOR EVALUATING PATIENT PROGRESS

An advantage of standardized scales such as the Beck Depression Inventory (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), Symptom Check List-90R (Derogatis, 1983), and the like is that a patient's score on one of these scales can be compared to a sample of "normals," the pretest scores of other patients, and the posttest scores of other patients, and compared against some absolute (widely accepted) standard of improvement. The ability to place all patients on a single continuum of distress eliminates most of the disadvantages of individualized scales while allowing for comparisons across therapists, treatment settings, and interventions.

Froyd, Lambert, and Froyd (1996) have identified the most frequently used scales reported in the literature on psychotherapy outcome. Many of these scales have excellent psychometric properties and an abundance of normative data, making them suitable outcome measures. Ogles, Lambert, and Masters (1996) have reviewed many of these instruments and provided evaluations of each, along with recommendations for their use by individual clinicians wishing to evaluate and track patient change.

In short, there is no shortage of suitable standardized scales for assessing patient outcome. They are readily available and easy to use.

An issue of considerable contemporary interest is the possibility of using such standardized scales in routine clinical practice by evaluating clinical change in relation to a cutoff score (national standard) for improvement. Since each patient that is treated evaluates his or her status on the same symptoms, comparisons between the effectiveness of clinicians, treatments, modalities, and the like can be made.

Based on the concept of social validation (Kazdin, 1977; Wolf, 1978) and statistically derived clinical significance (Jacobson, Follette, & Revenstorf, 1986; Jacobson and Truax, 1991; Jacobson, Wilson, & Tupper, 1988), methods have been developed to set standards for clinically meaningful patient change. The clinical significance methodology provides for the calculation of two specific statistical indexes (a cutoff point between normal and dysfunctional samples and an evaluation of the reliability of the change score) that provide a specific guideline for interpreting patient change.

The use of standardized scales and cutoff scores can be illustrated by reference to our own work. First, clinicians are asked to rate patient change on a session-by-session basis in much the same way Clement recommends, that is, ideographic assessment on the basis of specific patient problems. But, in addition, the patient is asked to complete a standard rating scale before each therapy session. Their progress on both individualized goals and the standardized scale is displayed graphically to the clinician immediately after the data are entered and these data are also available for broad-based studies of psychotherapeutic success.

The selection of a standardized outcome measure was a difficult decision. In the end, we chose the Outcome Questionnaire (OQ; Lambert, Lunnen, Umphress, Hansen, & Burlingame, 1994). This inventory was designed to measure three areas of patient functioning: symptomatic distress, interpersonal problems, and social role adjustment. It includes not only items assessing the intensity of symptoms but items measuring positive mental health or quality of life. This broad focus on patient functioning seemed more appealing than choosing a mono-symptomatic scale such as the Beck Depression Inventory (even though there is evidence to suggest that mono-

symptomatic scales do not merely measure the symptoms denoted by the scale name). The content of the OQ was consistent with the nature of symptoms found in a broad spectrum of Employee Assistance Program, outpatient, and inpatient samples. It had the following advantages: (a) brief (45 items) and therefore suitable for repeated measurement on a weekly basis, (b) solid psychometric properties, (c) sensitive to change over short time periods, and (d) inexpensive.

Once a suitable measure appropriate for a wide range of patients was selected, a cutoff point for demarking the dysfunctional range of functioning from the functional range was calculated. Normative functioning was identified by taking the scores of a sample of community volunteers (normals) and a patient sample (community mental health center outpatients) and applying the Jacobson et al. (1984) formula for clinically significant change cutoffs: $(SD_1M_2 + SD_2M_1)/(SD_1 + SD_2)$.

Reliable Change Index

The reliable change index (RC) was developed to assess the statistical reliability of the change a subject exhibits when he or she has crossed a cutoff point. If the RC is larger than 1.96, then it is with a 95% confidence level that reliable change can be said to have occurred (Jacobson & Truax, 1991). To determine whether a test score change is due to chance variation or to an actual change in the underlying characteristic, the formula $RC = (X_1 - X_2)/S_{diff}$ is used, where X_1 = a subject's pretest scores, X_2 = the same subject's posttest score, and S_{diff} = the standard error difference between the two test scores. S_{diff} describes the spread of the distribution of change scores (posttreatment minus pretreatment) that would be expected if no actual change had occurred and is defined as $2(SE)^2$, where SE = the standard error of measurement. SE is defined as $s_1(1 - r_{xx'})$, where s_1 = the standard deviation of the pretreatment sample and $r_{xx'}$ = the test-retest reliability of the measure.

For the normative samples under consideration, the cutoff on the OQ was determined to be a score of 67. Reliable change was determined to be 15 points. Therefore, when a patient moves from the dysfunctional distribution into the functional distribution by passing the cutoff of 67 and changes by 15 points or more, that person is judged to have had a clinically significant improvement.

To illustrate the use of this methodology, Figures 1–3 present the graphs of three patients whose improvement followed different courses. Each patient was assessed prior to therapy and before each session of therapy. In this particular study (Kadera, Lambert, & Andrews, 1995) neither the patient nor the therapist was aware of the patient's OQ score. Patient A (Figure 1), who was diagnosed with a major depression, was seen for 29 sessions. At the pretest, his OQ score was 84. At treatment termination, his OQ score was 29. The score passed the cutoff of 67 and changed by 55 points; thus, Patient A is an example of clinically significant improvement. Although the patient passed the cutoff for clinically significant change by the 11th session, the patient and therapist continued to work on specific goals and on relapse prevention until both agreed that therapy was no longer necessary.

The graph of Patient B (Figure 2) represents a different picture of change. She began therapy with an OQ score of 78 and terminated from therapy after session nine without discussing termination with the therapist. At this time Patient B was not improved; in fact, she had become more symptomatic, having an OQ score of 100. The therapist was disappointed with the outcome for this patient, who had a bipolar diagnosis. The patient would not comply with instructions to take medication. Patient B remained dysfunctional at a follow-up 6 months later (not represented in the graph).

Patient C (Figure 3) did not show a consistent pattern of change. He began therapy just within the dysfunctional range with a pretreatment OQ score of 70 and remained in the dysfunctional range until just before treatment termination. At termination Patient B had an OQ score of 66, had crossed the cutoff into the functional sample, but had not made reliable change. Both the therapist and patient were satisfied with the outcome of treatment, including goal attainment. The jump in distress occurring between the fourth and eighth sessions of therapy was apparently due to a family crisis that the patient discussed extensively in therapy.

The use of a rating scale with a standard definition of improvement has obvious advantages over the use of ideographic measures alone. However, several issues must be addressed in evaluating treatment outcome. What if a patient achieves his or her goals but fails to reach the standard for clinical significance on a rating scale (Patient C)? Or vice versa? How long must a patient remain in the

Subject A

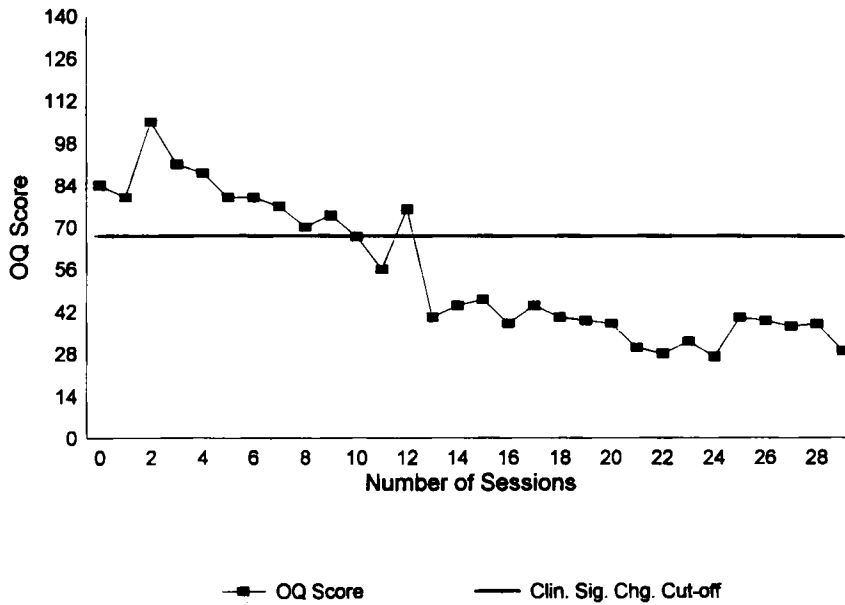


Figure 1. Weekly Outcome Questionnaire ratings of a patient undergoing treatment: Patient A, clinically significant change.

Subject B

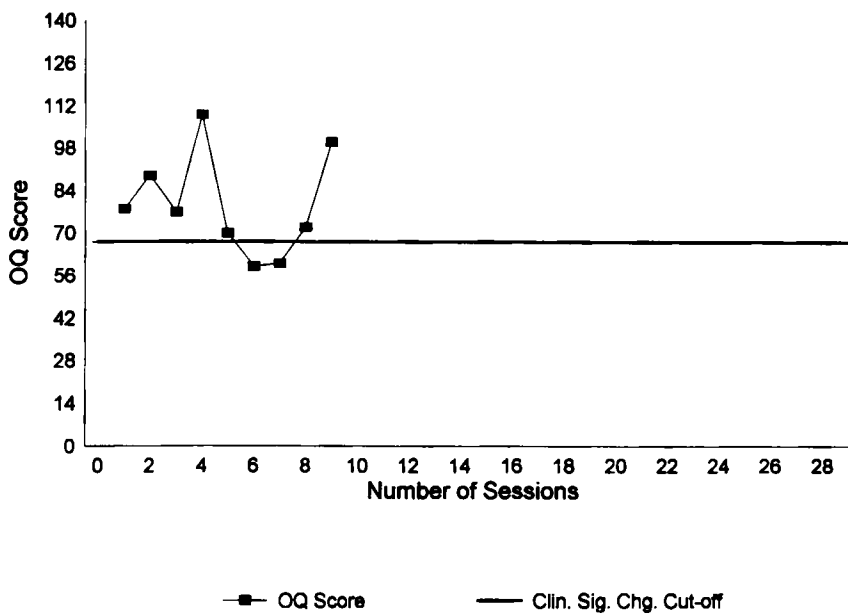


Figure 2. Weekly Outcome Questionnaire ratings of a patient undergoing treatment: Patient B, deterioration.

functional range to be confident that they have attained a reasonably stable or lasting improvement (Patient A)? What if a patient begins therapy in the functional range? Space does not permit a full discussion of these and other

issues, but routine solutions to these problems can be adopted. The point to be made here is that standardized scales can provide interesting data that can be used both clinically and for research purposes.

Subject C

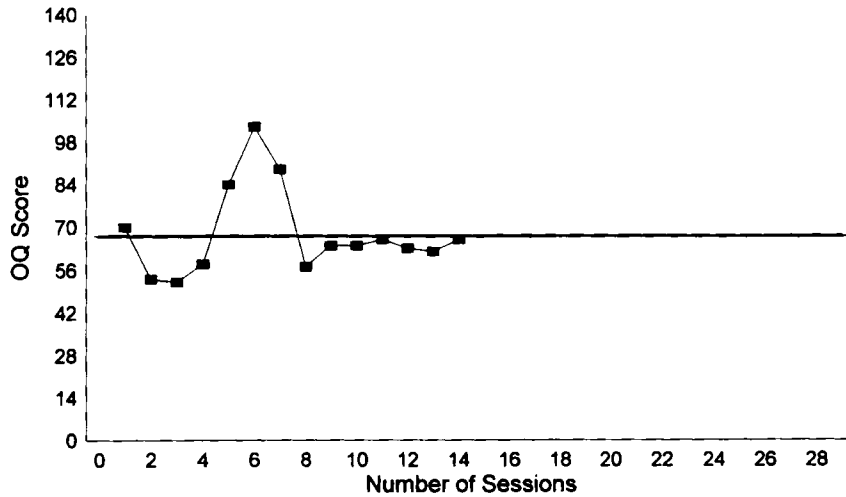


Figure 3. Weekly Outcome Questionnaire ratings of a patient undergoing treatment: Patient C, no change.

■ OQ Score — Clin. Sig. Chg. Cut-off

OUTCOME ASSESSMENT IN CLINICAL PRACTICE AND THE MANAGED CARE ENVIRONMENT

The behavioral health industry is increasingly evolving from a solo-practitioner delivery model into one dominated by managed care organizations (MCOs), which, at their best, seek to serve patients through organized systems of care designed to achieve optimal clinical outcomes in a cost-effective manner. Like individual practitioners, such systems must have the ability to monitor quality as well as the cost of care delivered and the clinical effect.

The wholesale purchasers of health care services (employers, insurance companies, government) are increasingly demanding that measurement of clinical outcomes be incorporated into all aspects of health care delivery, and that treatment decisions are made based on sound empirical evidence for efficacy and value rather than financial considerations alone. In response to this market demand, the major MCOs are moving rapidly to develop the methodology to incorporate the measurement of clinical outcomes into their delivery systems (Bartlett and Cohen, 1993; Brown, Fraser, & Bendoraitis, 1995). Advances in computing and communication technologies have now made it feasible to economically connect the individual practitioner with the information

systems of the MCOs. This next generation of information systems has come to be referred to as Clinical Information Systems (Brown et al., 1995).

Managed care companies are working collaboratively with software development firms to produce practice management software that will not only perform traditional practice management functions such as scheduling, accounting, and billing, but will also facilitate the collection of clinical data through the use of an electronic patient record. Alternatively, easy-to-use scannable forms are being developed that permit the provider to mail or fax the necessary data to the MCO. Individual clinicians would do well to employ the same systems in their clinical practice because these systems will be compatible with those used by large practitioner groups and because the data could contribute to the emerging information base that may be used to improve clinical practice.

A well-designed electronic patient record must incorporate the capacity to track ideographic data including individualized treatment goals, as well as nomothetic data such as standardized measures of patient impairment and improvement.

The MCO will be able to profile providers (and providers will be able to profile themselves) based on the clinical outcomes obtained (with appropriate case-mix

adjustment) as well as the cost of obtaining the outcome. This can only be achieved through the use of standardized assessment instruments that allow comparisons with other providers. The use of standardized scales will allow individual providers to have access to case-mix adjustment data and thereby allow comparisons of their case load with others who have a similar case load.

A further advantage of the use of standardized measures in such a system is the possibility of developing algorithms that can be incorporated into the MCOs' clinical information systems (Gray and Glazer, 1994). The use of such algorithms will permit the MCO to rely less on the expensive and intrusive inspect-and-control methods of having clinicians individually review each case, and increasingly rely on standardized, empirical data to monitor the care provided. Based on empirically derived indicators, cases can be targeted for review on an exception-only basis. This offers the potential to dramatically reduce the non-value-added work for both the practitioner and the MCO associated with reviewing cases that are in fact proceeding well. It will enable the individual provider to use the information provided by algorithms in their solo or group practice to enhance and standardize decision making.

Since outcome research persistently shows that treatment is effective (Lambert and Bergin, 1994), clinicians can welcome the emphasis on evaluation of treatment effects apparent in present-day mental health care. There is no doubt that standardized monitoring of patient progress will further demonstrate the value of psychotherapy and lead to greater understanding of just how and when it is most and least effective. We look forward to the time when clinicians routinely and systematically gather and share information about the effects of their work, blending the best of the art and science of psychotherapy.

NOTE

1. We might also note that Clement's recommendations regarding effect-size calculation need to be interpreted cautiously. Clement proposes the calculation of an effect size to express, in standard score units, the degree of patient progress. The reader who is not familiar with the effect-size statistic should be warned that the effect sizes obtained by following Clement's recommendations will be exceptionally large in comparison to those produced in controlled outcome studies. There are two reasons for this. First, effect sizes based on pre-to-post change (without taking into account changes that occur in no-

treatment or placebo controls) are often at least double those that are based on the procedures recommended by Smith, Glass, and Miller (1980). Second, effect sizes based on individualized outcome measures are typically much larger than those produced by standard rating scales. Thus, the application of Clement's proposal will result in exceptionally high effect sizes. These effect sizes cannot be meaningfully compared with effect sizes produced from standard rating scales, or added to them to obtain an average effect size. In short, the large effect sizes presented by Clement may be more a function of methodology than unusually successful therapy and should not be compared with effect sizes typically reported in the treatment outcome literature.

REFERENCES

- American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., revised). Washington, DC: Author.
- Bartlett, J., & Cohen, J. (1993). Building an accountable, improvable delivery system. *Administration and Policy in Mental Health, 21*(1), 51-58.
- Battle, C. C., Imber, S. D., Hoehn-Saric, R., Stone, A. R., Nash, E. H., & Frank, J. D. (1966). Target complaints as criteria for improvement. *American Journal of Psychotherapy, 20*, 184-192.
- 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.
- Brown, G. S., Fraser, J. B., & Bendoraitis, T. M. (1995). Transforming the future—the coming impact of CIS. *Behavioral Health Management, 14*(5), 8-12.
- Clement, P. W. (1996). Evaluation in private practice. *Clinical Psychology: Science and Practice, 3*, 146-159.
- Derogatis, L. R. (1983). *SCL-90: Administration, scoring, and procedures manual for the revised version*. Baltimore: Clinical Psychometric Research.
- Froyd, J. E., Lambert, M. J., & Froyd, J. D. (1996). A review of practices of psychotherapy outcome measurement. *Journal of Mental Health, 5*, 11-15.
- Gray, G. V., & Glazer, W. M. (1994, January/February). Psychiatric decision making in the 90's: The coming era of decision support. *Behavioral Healthcare Tomorrow*.
- Jacobson, N. S., Follette, W. C., & Revenstorff, D. (1986). Toward a standard definition of clinically significant change. *Behavior Therapy, 17*, 308-311.
- Jacobson, N. S., Follette, W. C., Revenstorff, D., Baucom, D. H., Hahlweg, K., & Margolin, G. (1984). Variability in outcome and clinical significance of behavioral marital therapy: A reanalysis of outcome data. *Journal of Consulting and Clinical Psychology, 52*, 497-504.

- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12–19.
- Jacobson, N. S., Wilson, L., & Tupper, C. (1988). The clinical significance of treatment gains resulting from exposure-based interventions for agoraphobia: A reanalysis of outcome data. *Behavior Therapy, 19*, 539–552.
- Kadera, S. W., Lambert, M. J., & Andrews, A. A. (1995, June). *How much therapy is really enough? A session-by-session analysis of the psychotherapy dose-effect relationship*. Paper presented at the annual meeting of the Society for Psychotherapy Research, Vancouver, BC, Canada.
- Kazdin, A. E. (1977). Assessing the clinical or applied importance of behavior change through social validation. *Behavior Modification, 1*, 427–452.
- Kiresuk, T. J., & Sherman, R. E. (1968). Goal Attainment scaling: A general method for evaluating comprehensive community mental health programs. *Community Mental Health Journal, 4*, 443–453.
- Lambert, M. J., Lunnen, K., Umphress, V., Hansen, N., & Burlingame, G. M. (1994). *Administration and scoring manual for the Outcome Questionnaire (OQ-45. 1)*. Salt Lake City: IHC Center for Behavioral Healthcare Efficacy.
- Lambert, M. J., & Bergin, A. E. (1994). The effectiveness of psychotherapy. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 143–189). New York: Wiley.
- Lambert, M. J., & Hill, C. E. (1994). Assessing Psychotherapy outcomes and processes. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 72–113). New York: Wiley.
- Marten, P. A., & Heimberg, R. G. (1995). Toward an integration of independent practice and clinical research. *Professional Psychology: Research and Practice, 26*, 48–53.
- Mintz, J., & Kiesler, D. J. (1982). Individualized measures of psychotherapy outcome. In P. C. Kendall & J. N. Butcher (Eds.), *Handbook of research methods in clinical psychology* (pp. 491–534). New York: Wiley.
- Ogles, B. M., Lambert, M. J., & Masters, K. (1996). *Assessing outcome in clinical practice*. New York: Allyn & Bacon.
- Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *The benefits of psychotherapy*. Baltimore: Johns Hopkins University Press.
- Waskow, I. E., & Parloff, M. B. (1975). *Psychotherapy change measures*. Rockville, MD: National Institute of Mental Health.
- Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis, 11*, 203–214.

Received July 12, 1995; accepted July 12, 1995.