Mental Health Services Research, Vol. 2, No. 2, 2000

Measurement of Fidelity in Psychiatric Rehabilitation

Gary R. Bond,[1,2] Lisa Evans,[1] Michelle P. Salyers,[1] Jane Williams,[1] and Hea-Won Kim[1]

Until recently, most psychiatric rehabilitation models have been poorly defined and few have had systematic methods for measuring their implementation. We review the historical roots for the development of fidelity measures and describe recent applications in both research and practice.

KEY WORDS: fidelity; program implementation; program standards; program models; psychiatric rehabilitation.

Fidelity refers to the degree to which a particular program follows a *program model.* By program model we mean a well-defined set of interventions and procedures to help individuals achieve some desired goal. *Fidelity measures* are tools to assess the adequacy of implementation of program models. Increasingly, measures of program model fidelity have become standard requirements in mental health services research (Heflinger, 1996; Henggeler, Pickrel, & Brondino, 1999) and in other applied fields, such as criminal justice, education, and medicine (Leithwood & Montgomery, 1980; Rezmovic, 1984; Schreier & Rezmovic, 1983). Despite this attention, no systematic body of theory and research on fidelity has yet appeared in the mental health services area.

In this paper we review the promise of fidelity measures for advancing the research and practice of one area of mental health services, namely, psychiatric rehabilitation. By *psychiatric rehabilitation* we mean services and programs intended to help adults with severe mental illness attain optimal integration into normal adult roles in the community.

Psychiatric rehabilitation has evolved in an unusually eclectic and pragmatic fashion because many areas lack clearly articulated models or theory explicitly linked to the nature of severe mental illness (Hogarty, 1995). In addition, many influential leaders in the psychiatric rehabilitation field have emphasized the advantages of eclecticism, innovation, and experimentation in program design (Dincin, 1995). Psychiatric rehabilitation approaches typically are classified according to areas of role functioning, namely community integration, independent living, employment, academic achievement, social relationships, communication skills, and family relationships (Dincin, 1995). Corresponding to this list are the following broad categories of psychiatric rehabilitation approaches: *case management* (community integration), *residential programs* (independent living), *vocational programs* (employment), *supported education* (academic achievement), *drop-in centers* (social relationships), *skills training* (communication skills), and *family psychoeducation* (family relationships). Within each of these categories are specific program models. Models of case management, for example, include assertive community treatment, the strengths model, the rehabilitation model, and the brokered model (Solomon, 1992).

In this paper we provide a historical context for the development of fidelity measurement. We then offer examples of the ways that fidelity measures have been and could be used. We examine two separate uses and traditions—the scientific and the practical.

[1]Department of Psychology, Indiana University-Purdue University Indianapolis, Indianapolis, Indiana.

[2]Correspondence should be directed to Gary R. Bond, Department of Psychology, Indiana University-Purdue University Indianapolis, 402 North Blackford Street, Indianapolis, Indianapolis 46202-3275; e-mail: gbond@iupui.edu.

ORIGINS OF FIDELITY MEASURES IN RESEARCH APPLICATIONS

The concept of fidelity emerged in the 1960s, when psychotherapy researchers discovered the im-

1522-3434/00/0600-0075$18.00/0

possibility of sorting out the methodological and interpretative problems in the early outcome studies (Moncher & Prinz, 1991; Waltz, Addis, Koerner, & Jacobson, 1993). Researchers naively assumed that therapists subscribing to particular forms of psychotherapy intervened in a consistent and distinctive fashion based on their affiliation (e.g., client-centered, rational-emotive, psychodynamic). Eysenck (1952) and other critics pointed out the weaknesses in these arguments. In seeking to refute critics, it became increasingly clear that psychotherapy models were not operationally defined. Replication studies were difficult, because most studies did not provide sufficient information on the treatment used. It also became increasingly obvious that psychotherapy approaches with different labels often overlapped. Moreover, even when models *were* dramatically different conceptually, there was often little assurance that the therapists representing different approaches would show this behaviorally. In fact, studies showed that therapists subscribing to a particular approach did not necessarily behave similarly, nor did therapists subscribing to theoretically very different approaches always behave that differently (Lieberman, Yalom, & Miles, 1973; Sloane, Staples, Cristol, Yorkston, & Whipple, 1975).

Client-centered therapy was among the first to systematically examine its therapeutic methods and techniques (Rogers, 1951, 1957). In an early formulation, Rogers defined the three critical ingredients of client-centered therapy: *unconditional positive regard, genuineness,* and *empathy,* which he hypothesized to be the ''necessary and sufficient ingredients of effective psychotherapy.'' These constructs were used as the basis for the development of process ratings scales—what we would call fidelity measures. The formulation of the critical ingredients of client-centered therapy, along with empirical methods to measure them, spawned an enormous and highly productive literature, which has helped to shape later theoretical and practical formulations of basic listening techniques in psychotherapy (Mitchell, Bozarth, & Krauft, 1977). A similar process is unfolding in the psychiatric rehabilitation field in relation to a few well-defined and experimentally evaluated program models.

Fidelity measurement accelerated the maturation of psychotherapy research by making standardized treatments possible and by providing methods to document differences between different forms of treatment. The measurement of fidelity developed in two directions associated with two related methodological issues (Moncher & Prinz, 1991). The first, referred to as *treatment integrity,* concerns the degree to which a treatment condition is implemented as intended. The second, referred to as *treatment differentiation,* refers to ''whether treatment conditions differ from one another in the intended manner such that the manipulation of the independent variable occurred as planned'' (Moncher & Prinz, 1991, p. 247). In a similar vein, Waltz *et al.* (1993) suggested four types of therapist behaviors to be considered in assessing adherence: (a) behaviors that are unique and essential to the model; (b) behaviors that are essential but not unique; (c) behaviors that are compatible with the model, but are neither necessary nor unique; (d) behaviors that are prohibited. A comprehensive list of essential and prohibited behaviors for a model provides a blueprint for assessing treatment integrity and for differentiating a program model from others.

The twin concepts of model integrity and model differentiation are directly applicable to the mental health services field, where program labels abound and where the same labels often are used with different meanings. To take one such example, *intensive case management, assertive case management, mobile treatment teams, continuous treatment teams,* and *assertive community treatment* are sometimes intended to mean the same thing, sometimes something different (Marshall & Creed, in press). Although experts may agree to some extent on what is meant by each of these program labels (McGrew & Bond, 1995; Schaedle & Epstein, 2000), there is no unanimity on elements for these models that are unique, essential, compatible, and prohibited. Behaviors that are essential in one program model but prohibited in another are particularly useful for treatment differentiation. For example, one study contrasted two vocational approaches which were sharply differentiated on the job placement process (initial prevocational skills training versus rapid job search) and on the location of services (at a separate rehabilitation agency versus at the mental health center) (Drake, McHugo, Becker, Anthony, & Clark, 1996).

In the psychotherapy literature, the next logical step after developing measures for assessing adherence to a program model was to tease out *which* ingredients had the most influence on client outcomes. The search for critical ingredients led to a broader inquiry into ''process'' variables influencing psychotherapy outcomes. The ensuing explosion of psychotherapy process-outcome research has been staggering. Orlinsky, Grawe, and Parks (1994) identi-

fied 2354 separate findings in studies conducted over a three-decade period.

The notion of operationally defining program models had another set of implications for research, namely, that elements of an effective program model could be systematically implemented by training staff in the application of the model principles and techniques. Following the client-centered therapy example, the standard in psychotherapy has been to develop ''treatment manuals'' (also called practice manuals), which provide detailed descriptions of how a program should be organized and how providers should perform their responsibilities.

Unfortunately, psychotherapy researchers have not always employed even minimal efforts at establishing fidelity. In a review of 359 psychotherapy studies, Moncher and Prinz (1991) found that 32% used a treatment manual, 22% had a systematic method for ensuring supervision of therapists, and 18% checked adherence to the protocol. Altogether, only 6% of the studies in their review employed all three methods for enhancing fidelity, and 55% did not use any of these measures.

Fidelity in Psychiatric Rehabilitation

Like the early psychotherapy literature, much of the psychiatric rehabilitation literature has lacked rigorous assessment of adherence to program models. In his review of 33 controlled studies of community support programs (i.e., psychiatric rehabilitation), Brekke (1988) found that only one satisfied his criteria for a complete program description. More recently, a case management review found that 79% of the studies failed to procedurally define program models (Gorey *et al.,* 1998). Other case management reviews also observe the lack of program model clarity (Latimer, 1999b; Marshall & Creed, in press; Marshall, Lockwood, Green, & Gray, 1998; Mueser, Bond, Drake, & Resnick, 1998). Reviews of vocational rehabilitation for people with severe mental illness similarly have noted the ambiguities in program descriptions (Bond, Drake, Becker, & Mueser, 1999a). In fact, in all of the psychiatric rehabilitation areas noted above, commentators have made similar observations about model ambiguity and lack of adequate program descriptions.

Measurement of fidelity in psychiatric rehabilitation has not been completely ignored, however. In a pioneering effort, Paul and his colleagues developed a hospital-based social learning approach with an intricate method for assessing program fidelity (Paul & Lentz, 1977). Their measurement approach was not widely adopted, partly because of its labor-intensive requirements. In another early effort to measure fidelity, Anthony, Cohen, and Farkas (1982) identified 10 ''essential ingredients'' for determining if a program followed psychiatric rehabilitation principles. They later developed a formal coding system for assessing partial hospitalization programs on these principles (Fishbein, 1988). Anthony and his colleagues then developed a series of training manuals to provide concrete guidelines for implementing their approach (Anthony, Cohen, & Pierce, 1980). Their approach was not intended to define a specific model, but rather a set of practices that are applied in different practice settings (Anthony, 1994).

The psychiatric rehabilitation model with probably the longest history of attention to model standards is the assertive community treatment (ACT) model, developed by Stein and Test (1980). Their landmark study was important not only for demonstrating an effective alternative to hospitalization, but also for the clarity of its program model. Their work laid the foundation for a process study of ACT (Brekke & Test, 1987) and the later development of ACT fidelity scales (e.g., McGrew, Bond, Dietzen, & Salyers, 1994; Teague, Bond, & Drake, 1998; Teague, Drake, & Ackerson, 1995).

Advances in Measurement

Although paper-and-pencil surveys of staff and program participants show some promise for differentiating among program models (Burt, Duke, & Hargreaves, 1998; Jerrell & Ridgely, 1999), most observers would argue for the importance of using a multimodal approach, including chart review, observation of team meetings, service logs, etc., in addition to surveys (Hargreaves, Shumway, Hu, & Cuffel, 1998). Using a diversity of data sources, Brekke showed that systematic measurement of theoretically relevant dimensions of psychiatric program models was possible (Brekke, 1987; Brekke & Aisley, 1990; Brekke & Test, 1992; Brekke & Wolkon, 1988). His work highlighted the feasibility of model differentiation through empirical methods.

Progress in developing psychiatric rehabilitation fidelity measures has been hampered by several factors. One major factor has been the lack of well-defined models. With the exception of ACT (Allness & Knoedler, 1998; Stein & Santos, 1998), skills

training (Wallace, Liberman, MacKain, Blackwell, & Eckman, 1992), the Individual Placement and Support (IPS) model of supported employment (Becker & Drake, 1993), and some family approaches (Mueser & Glynn, 1999), practice manuals have been rare in psychiatric rehabilitation. The development of fidelity scales is much easier with detailed practice manuals.

The complexity of psychiatric rehabilitation services poses a great challenge to fidelity measurement. Whereas in psychotherapy, the focus is on therapist behaviors, model fidelity in psychiatric rehabilitation typically concerns not only practitioner behavior, but also structural aspects of a program (e.g., caseload size, staff qualifications), location of services (e.g., in community settings), and "behind the scenes" activities (e.g., integration of treatment and rehabilitation). Manualizing counseling and psychotherapy often involves minute-to-minute specification of specific therapist interventions, whereas practice manuals for psychiatric rehabilitation models inevitably must be conceptualized at a more macro level. Thus, psychiatric rehabilitation models are inherently difficult to manualize because the interventions occur in multiple settings, with multiple providers and recipients, and involve diverse activities that go far beyond a counseling setting.

THE PROMISE OF FIDELITY MEASURES: RESEARCH APPLICATIONS

Fidelity measures serve many research purposes, including uses related to internal, discriminant, external, construct, and predictive validity (Moncher & Prinz, 1991). We distinguish four uses of fidelity, each addressing different aspects of validity: (a) ensuring model adherence in program evaluations, (b) facilitating communication in the literature, (c) synthesizing a body of research, and (d) identifying critical ingredients of program models. We illustrate each of these uses with examples from the psychiatric rehabilitation literature.

Ensuring Model Adherence in Program Evaluations

Moncher and Prinz (1991) define fidelity of treatment in outcome research as "confirmation that the manipulation of the independent variable occurred as planned" (p. 247). Measuring program fidelity, then, can be viewed as a "manipulation check," intended to determine if the independent variable yielded the desired difference between the treatment groups. However, unlike classic experimental studies, in which the researcher manipulates one specific variable while holding all other variables constant, studies evaluating the effectiveness of a program model usually must contend with comparisons between two or more treatments that differ on many dimensions.

In order to obtain significant outcome differences, the designs for outcome studies should "maximize the systematic variance" (Kerlinger, 1986) by ensuring that the experimental and control conditions are different. Logically, there are two strategies for making the conditions different. One is by employing a neutral control condition (i.e., one that is not expected to result in large improvements in outcomes), and the other is to use a powerful experimental condition. (Of course, a control condition that represents standard services will be of more scientific and practical interest than an approach already known to be ineffective.) Making sure the treatments are different is one main way to increase statistical power (Lipsey, 1990). More generally speaking, it is important to examine the implementation of both the experimental and control groups along the same study dimensions in order to determine the degree of treatment differentiation, which is the systematic variance that is expected to account for any differences in outcomes.

Several multisite randomized controlled trials illustrate the use of fidelity measures to assess model adherence. Fidelity of program implementation was examined in a New Hampshire multisite study comparing ACT to standard case management (Teague *et al.,* 1995). Although demonstrating significant differences between experimental and control conditions, this study also found site differences, with some control programs rating relatively high on the ACT fidelity scale. In a subsequent analysis, McHugo, Drake, Teague, and Xie (1999) found that among the ACT clients in their sample, those in high-fidelity ACT programs had higher rates of retention in treatment, greater remission from substance use disorders, and fewer hospital admissions than those in low-fidelity programs. In an Indiana multisite ACT study, the investigators rated program fidelity after the fact, speculating that drift from the ACT model (e.g., lack of stringent admission criteria, increased caseload sizes, diffusion of staff responsibilities, and combined team meetings for ACT and standard case management) accounted for the lack of experimental differ-

ences in client outcomes (Bond, Miller, Krumwied, & Ward, 1988). In a vocational study, fidelity to a supported employment model was stronger in one of two sites, which also had more favorable employment outcomes (Drake *et al.,* 1996).

As the first example suggests, fidelity measures should be used to assess both experimental and control conditions. On the experimental side, the question is whether the program has been implemented as intended, or if it in actuality looks similar to habitual practices of service provision. On the control side, the question is typically whether the control group has been contaminated by imitating the strategies of the experimental program. In the New Hampshire ACT study, the main source of infidelity was that standard case management drifted toward the ACT model; in the Indiana study, the problem was the opposite, namely that the newly developed ACT team at one site did not fully embrace the model. In the latter case, a lack of staff training and program monitoring may have been a major factor. Such program drift is unfortunately quite common in the literature (Bond, 1991; Rosenheck, Neale, Leaf, Milstein, & Frisman, 1995). Rather than simply documenting failures to implement, a far more useful approach is to use fidelity measures in conjunction with systematic efforts to achieve excellence in program implementation, to help guide development and keep programs on course (Henggeler *et al.,* 1999). Finally, as these examples also suggest, the measurement of fidelity is especially important in multisite studies, in which findings may be stronger in some sites than in others.

A key issue in ensuring program fidelity is to make sure that the program model is clearly defined at the outset. Initial ambiguity about the program model may have been a factor in the ambiguous findings in two recent large-scale studies (Becker, Holloway, McCrone, & Thornicroft, 1998; Burns *et al.,* 1999). In both studies, intensive case management was defined largely in terms of lower caseload ratios without detailed prescriptions for program implementation.

Another fidelity issue that is not frequently discussed in the literature is the need for *continued* fidelity measurement over the course of a study, especially for multiyear studies. Programs change over time, sometimes showing marked improvement as they move beyond their startup period. Conversely, they sometimes lose their wholehearted commitment to the program model, for example, as they approach the end of grant funding, especially if this means (as it often does) that staff workers are transferred to other positions (McHugo *et al.,* 1998).

Fidelity measures are equally valuable when program evaluations do yield significant treatment outcomes. In addition to increasing the confidence in the study's internal validity, the fidelity measures provide a roadmap for replication. Moreover, with significant findings, the research questions shift to asking what the critical ingredients of program success are. Fidelity measures provide the basis for more fine-grained inquiry into such questions.

Facilitating Communication in the Literature

One source of confusion in the psychiatric rehabilitation literature has been distinguishing between related program approaches within specific domains. Earlier, we gave the example of the case management area, in which many different variants have been reported in the literature. A recurring question has been the conceptual and empirical overlap between these different models (Mueser *et al.,* 1998).

The Dartmouth ACT Scale (DACTS) is an example of a fidelity scale that has been helpful in mapping out a psychiatric rehabilitation domain (Teague *et al.,* 1998). Although developed to discriminate well-implemented ACT programs from traditional (less intensive) case management services, the DACTS also may be useful for delineating a typology of case management services in general. Teague *et al.* (1998) piloted the DACTS in 50 case management programs representing four distinct types of service models: ACT, intensive case management provided by the Veterans Administration, outreach programs for people who were homeless and mentally ill, and traditional case management. The DACTS discriminated across the four types of case management, consistent with predicted order of similarity to ACT. However, a subsequent study comparing DACTS ratings in a sample of 18 outreach programs for homeless people with mental illness found nonsignificant differences between programs subscribing to nominally different program models (Johnsen *et al.,* 1999). Thus, more work is needed to determine the capacity of the DACTS to make fine-grained discriminations.

Another example of the use of fidelity measures for mapping out a domain is given by Bond, Becker, Drake, and Vogler (1997), who found that vocational programs subscribing to the IPS model of supported employment sharply differed from "traditional" vocational services across a wide range of observable

criteria. The IPS fidelity scale also found less dramatic, but significant differences between IPS and other forms of supported employment.

One variation on the theme of using fidelity scales to map out a domain is *defining model adaptations.* Once a program model is well defined, variations of it can be assessed in relation to the original model. An example illustrating the use of a fidelity scale to measure model adaptation is given in a study examining the effectiveness of transferring clients from an ACT program to a modified ACT program referred to as a ''step-down'' program (Salyers, Masterton, Fekete, Picone, & Bond, 1998). This step-down program emulated many of the positive features of ACT, while serving clients at a less intensive level. The ratings on the DACTS indicated specific ways in which the step-down program differed from the parent ACT program (e.g., frequency of team meetings, percentage of home visits) and ways in which it was similar (e.g., use of multidisciplinary team, focus on practical problems).

Synthesizing a Body of Research

Literature reviews aim at understanding the extent to which findings from individual studies are generalizable. In integrating the literature on the outcomes for a program model (or for an entire service domain, such as vocational services), reviewers face the dilemma of determining which studies to include and how to weight those that are included. Many factors go into these decisions, but one main consideration is the fidelity of implementation. The ideal circumstances for a reviewer would be a body of studies in which all investigators prospectively applied a single standardized fidelity measure which previously had been shown to be psychometrically adequate. Under these ideal circumstances, the reviewer could then establish a minimum criterion for program fidelity. If a program fell below the criterion, then that study would be excluded from the review. Alternatively, fidelity scores could be used as an independent variable in a meta-analysis (Lipsey, 1990).

Unfortunately, there are no examples in the mental health services area that come anywhere close to this ideal. Few domains have a critical mass of studies, and even fewer have used prospective fidelity ratings that would lend themselves to this procedure. However, two reviews using retrospective fidelity ratings hint at the promise this methodology might have. In one synthesis, McGrew *et al.* (1994) retrospectively coded 18 programs on a fidelity index which was correlated with a program-level client outcome measure. This study found a strong correlation between the fidelity index and reduction in hospital use. More recently, Latimer (1999b) retrospectively coded programs within a sample of 34 ACT studies using a simplified fidelity scale and found that high-fidelity ACT programs reduced hospital days more than low-fidelity programs.

Identifying Critical Ingredients

A fourth use of fidelity scales is to help identify critical ingredients that predict client outcomes. Critical ingredients refer to the elements of a model, such as the caseload ratio or location of services, which account for its effectiveness. In this application, theoretically important ingredients are represented by items or subscales on a fidelity measure. The usual method for demonstrating empirically that a program element is a *critical* ingredient is by obtaining a significant correlation with a criterion measure (i.e., a measure of client outcome). Following the logic of this design, the criterion measures should be congruent with the purposes of the program model; for example, a vocational model should have a primary impact in the employment domain. Although research has used different statistical methods, one typical strategy is to convert individual client outcomes into aggregate program-level measures (e.g., percentage employed). This general strategy relates to the predictive validity of a fidelity measure, by examining its relationship to outcome, and to construct validity, by showing the contribution of the elements of the model to program effectiveness.

A study illustrating this strategy was conducted by McGrew *et al.* (1994). The study was based on a 17-item scale, the Index of Fidelity to ACT (IF-ACT), which assessed objective features of a program, such as the inclusion of a nurse on the team, frequency of team meetings, and frequency of contacts with clients in the community. The items were retrospectively coded for 18 ACT programs in completed studies. Five of the 17 fidelity items significantly predicted reduction of hospital use: shared caseloads, total number of contacts, 24-hour availability, nurse on the team, and daily team meetings. Thus, this study suggested that specific program ingredients were associated with better client outcomes.

We add some cautionary notes about the limits

of this strategy. One limitation is that the criterion-oriented method for identifying critical ingredients only makes sense if one is confident in the validity of the criterion outcome measure and the potency of the model for impacting that outcome. Thus, the reduction of hospital use is, by all accounts, an excellent criterion for ACT programs. Similarly, one might presume that the placement rate into employment would be a sensible criterion for vocational programs. However, defining the annual employment placement rate for a vocational program is not as simple a matter as it might seem. Complications include time frames, caseload mix issues, client status (open cases, program dropouts, and successful closures), and variability in job tenure. Such complications may have partly accounted for the disappointing findings in a study seeking to identify the critical ingredients of supported employment (Bond, Picone, Mauer, Fishbein, & Stout, 1999).

A second obvious limitation is that this method works poorly for small samples, especially if there is a restriction in the range of programs on either the predictor or criterion measure (Zahrt, Bond, Salyers, & Teague, 1999). From a practical standpoint, assembling large samples of sites with requisite fidelity and outcome measures is a formidable undertaking.

We earlier mentioned the evolution of the psychotherapy literature toward the quest for identification of general process measures that predict better outcomes, regardless of the therapy model used. A parallel development in the psychiatric rehabilitation field has been the development of general-purpose instruments that seek to measure program features across a broad range of models (Burt *et al.,* 1998; Jerrell & Hargreaves, 1991; Moos, 1974). Although these instruments have demonstrated their utility in a growing body of studies, the complexities involved make this undertaking quite ambitious.

ORIGINS OF FIDELITY MEASURES IN PRACTICAL APPLICATIONS

Although fidelity measures only recently have begun to receive attention in the psychiatric rehabilitation literature, efforts to enhance program quality through specification of criteria for best practices in psychiatric rehabilitation have been an enduring concern of program administrators and planners. For example, the National Institute of Mental Health sponsored a series of conferences in the 1970s culminating in a set of guidelines intended to help reform the practice in community mental health centers (CMHCs). A program following these guidelines became known as a community support program (CSP; Turner & TenHoor, 1978). CSP guidelines were intended to provide criteria for CMHCs to be more responsive to serving clients with severe mental illness. Over the past two decades, the changes in methods for financing public mental health also have directly influenced the movement toward defining psychiatric services. In the early days of deinstitutionalization, CMHCs were funded primarily through state general funds, which typically provided grants or contract payments for approved costs, with few linkages to performance measures. In the 1980s, state departments of mental health increasingly began using Medicaid as a source for funding mental health services, including case management and rehabilitation services. Medicaid reimbursed for specific rehabilitation and treatment services defined under the state Medicaid plan; hence, CMHCs increasingly have been required to be explicit in documenting what interventions they provided, to whom, and for how long (Hogan, 1999). The need to specify the components of specific levels of service provision (e.g., levels of case management) provides a natural bridge to the development of fidelity measures.

Recently, the medical field has witnessed the proliferation of *practice guidelines.* Like treatment manuals, practice guidelines explain what services to provide, to whom, and how. Unlike treatment manuals, practice guidelines often are not ''model specific,'' but rather indicate recommended services across a range of program models all targeted to a specific population (e.g., people with schizophrenia). Thus, expert panels formed by the American Psychiatric Association (Herz *et al.,* 1997) and other groups (McEvoy, Scheifler, & Frances, 1999) have developed clinical practice guidelines for the treatment of schizophrenia, including recommendations for both psychopharmacological and psychosocial interventions. Practice guidelines have been produced by professional organizations, such as the International Association of Psychosocial Rehabilitation Services (IAPSRS, 1997b) and the National Association of Case Management (Giesler & Hodge, 1998). Government agencies have also been active in commissioning task forces to generate guidelines (Cochrane, Durbin, & Goering, 1997; Latimer, 1999a; Torrey & Wyzik, 1997). Managed care organizations are also involved in the development and use of practice guidelines (Dewan & Carpenter, 1997).

An especially influential example of an outline for practice guidelines comes from the Schizophrenia Patient Outcomes Research Team (PORT) treatment recommendations, which consisted of 30 medication and psychosocial treatment guidelines (Lehman, Steinwachs, & PORT Co-Investigators, 1998). For example, the PORT recommends ACT services for persons with schizophrenia who are either at high risk for rehospitalization or heavy service users.

Thus, these trends in the medical field suggest a growing emphasis on concretely specifying what services should be provided to a target population. Because of the dearth of empirical studies, most of the guidelines produced to date pertaining to psychiatric rehabilitation have been based on ''clinical consensus'' rather than empirical studies, and some guidelines are statements of broad principle rather than useable, specific prescriptions for intervention. Putting this into perspective, Hughes (1999) noted that ''only 20% of all medical treatments meet the level of 'evidence-based' treatments when the requirement is a body of studies using a rigorous experimental design'' (pp. 11–12). Consequently, fidelity measures could be seen as serving a complementary function of spelling out the interventions and models described in the practice guidelines, as well as providing one tool needed for establishing a firmer empirical basis for guidelines.

Another influence on the development of fidelity measures as practical tools has been the desire of proponents of well-established approaches to maintain high standards as their model is disseminated. An example is given by the Fountain House clubhouse model (Beard, Propst, & Malamud, 1982). Although the roots of this model date to the 1940s, it has evolved over the years and, like other psychiatric rehabilitation models, has yielded many variants. Recently, a set of clubhouse standards was adopted by an international conference of clubhouses, partly as a reaction to model diffusion (Propst, 1992). These standards have been the basis for a certification process, which is operated by the International Center for Clubhouse Development (ICCD), located at Fountain House. The formal process of certification is based on a site visit by a group of approved clubhouse trainers, who use a semistructured guide in determining adherence to standards (Gold Award, 1999; Moxley, 1993). Clubhouses are classified into ''certified'' programs (i.e., those approved by the ICCD) and noncertified programs. Recently, a content analysis of ICCD site visit reports was used as a way to infer critical program ingredients, by demonstrating ways in which programs often fall short of certification standards (Wang, Macias, & Jackson, 1999). The technique of ''program model errors'' is reminiscent of a scale developed to identify psychotherapeutic errors (Suh, O'Malley, & Strupp, 1986). It is possible that a focus on areas in which programs typically deviate from a model—i.e., ''prohibited behaviors'' (Waltz *et al.,* 1993)—may prove to be an especially efficient way to measure fidelity.

Still another trend increasing the interest in fidelity measures has been the role of consumer organizations in advocating for quality services. In state-by-state surveys, Torrey and his colleagues rated the adequacy of mental health services for people with serious mental illness in the U. S., broadly characterizing such domains as case management and rehabilitation services in each state (Torrey, Erdman, Wolfe, & Flynn, 1990). The resulting state ranking provided a ''report card'' that consumers and families could use to evaluate and select among providers and to advocate for better services. The National Alliance of the Mentally Ill has used a similar checklist approach to rate managed care organizations (Hall, Edgar, & Flynn, 1997). These evaluations suggest potential practical uses of fidelity measures for consumers. It remains to be seen how well the instruments developed by researchers can be adapted to such purposes. For example, many consumer report cards emphasize customer satisfaction dimensions, e.g., promptness of response and friendliness of staff, which comprise a very different set of concerns than researcher-generated fidelity measures.

THE PROMISE OF FIDELITY MEASURES: PRACTICAL USES

Practical uses of fidelity measures can be conceptualized as a four-dimensional cube consisting of (a) purpose, (b) sample, (c) timing, and (d) target audience. With regard to purpose, fidelity measures can be used to communicate standards, monitor progress over time, identify outliers, compare programs to norms, or document the relationship between model adherence and outcome. With regard to sample, fidelity measures can be used with a single program or a sample of programs. Obviously, the number of programs involved influences the logistics of fidelity assessment. With regard to timing, fidelity measures can be introduced before an organization has even decided what program models might be implemented, at the initial stages of development of a

new program, or at any stage after implementation. Fidelity measurement can take place once or at multiple time points. With regard to target audience, many stakeholder groups, including funding agencies, program managers and other program staff, consumers, and their families, are interested in maintaining standards for the sake of attaining quality of care. Practical applications of fidelity measures can of course incorporate several purposes, samples, points in time, and target audiences.

Thus, the potential practical applications of fidelity measures in psychiatric rehabilitation are enormous. Many of these applications have already been attempted, although a complete inventory of actual applications has never been compiled. The intense interest in learning about and obtaining user-friendly fidelity checklists became apparent to us with the surprisingly enthusiastic response to an early publication of a fidelity instrument (McGrew *et al.,* 1994). When fidelity is used in a practical setting, the results often go unrecognized in unpublished reports and documents. ''Underground,'' however, we believe there are many groups who are using such tools. We illustrate practical applications of fidelity measures with the following examples.

Guidelines for Considering Adoption of a Model

One practical use of fidelity scales is to introduce a program model to groups who have not had firsthand experience with it. For example, if a state mental health authority is seeking to introduce a new set of psychiatric rehabilitation services, it is helpful if decision-makers have concrete details before adopting a specific program model. Fidelity scales can provide a template for thinking about practice guidelines, whether or not a specific program model is adopted as is. Examples of task forces to develop practice guidelines can be found throughout the United States (Barton, 1997; Torrey & Wyzik, 1997), Canada (Cochrane *et al.,* 1997; Latimer, 1999a), and overseas (Marshall & Creed, in press). In addition to practice guidelines, fidelity measures can be used as a quick reference guide to program design and as a starting point for estimating program costs.

Monitoring Programs

The history of deinstitutionalization has been a recurring cycle of reform movements, each one hoped to be the innovation that would take hold. Often with great fanfare, states announce initiatives to improve services, often through the introduction of a new program model. New York's early experience with supported employment is illustrative of the implementation problems that occur as provider agencies implement wildly disparate services, often falling short of the effectiveness promised by the initiative (Noble, 1991).

Currently a number of states and local authorities are using checklists (i.e., fidelity measures) as tools to help avert such mistakes. With appropriate databases, evaluators can provide cross-site monitoring of program implementation, making comparisons such as (a) between target programs and established norms (if they exist), (b) across regions of the state (e.g., rural versus urban), (c) between individual sites and state averages, and (d) within programs and groups of programs over time. The comparisons can help identify specific areas in which the state as a whole falls short of established norms; regional differences that may be reflective of varying populations, resources, local traditions, or other factors; individual sites that may be exceptionally well-implemented and worthy of recognition; individual sites departing from the intended model; and improvement over time as programs develop. From a management standpoint, it is valuable to know what sites are outliers, so that one can intervene early. Even a fairly crude fidelity measure may be capable of serving as an early warning system for such sites.

Program monitoring has been most widely used for ACT programs. Since the early 1980s, Michigan has sponsored the dissemination of ACT programs throughout the state, requiring new ACT teams to follow standards for program operation (Mowbray, Plum, & Masterton, 1998). Other states have followed suit (Deci, Santos, Hiott, Schoenwald, & Dias, 1995). Since 1996, Illinois has been monitoring agencies funded through their statewide ACT initiative (Zahrt *et al.,* 1999). Using the staff from the state mental health authority to make ratings, program planners have found the DACTS to be a useful tool for communicating program expectations and ensuring their implementation.

One variation of a statewide monitoring approach concerns converting an existing program to a new program model. For example, Rhode Island recently has been involved in converting their day treatment services to supported employment (McCarthy, Thompson, & Olson, 1998). State planners

used the IPS fidelity scale to help shape expectations for provider agencies.

One of the most ambitious efforts in measuring program fidelity at a statewide level has been undertaken in Kansas. With the help of the University of Kansas School of Social Welfare, community mental health centers are currently using a comprehensive packet of materials known as "Best Practices Fidelity Tools" (Rapp, 1999). This packet spans the important domains of mental health services for adults with severe mental illness, providing specific behavioral indicators of what programs should be achieving in each service domain. Another example is given by Connect98, an initiative of the Illinois Office of Mental Health. This statewide program is using a set of fidelity tools to assess implementation of three psychiatric rehabilitation components: peer support, vocational services, and skills training (Bond, Evans, Kim, & Goodman, 1999b).

DISCUSSION

We are seeing increasing demands for the measurement of adherence to program standards, not only from the scientific community, but also from a variety of stakeholders involved in funding, providing, and receiving psychiatric rehabilitation services. Increasingly, journal editors are insisting that empirical studies include fidelity measures. Review panels for grants submitted to the National Institute of Mental Health and to other federal agencies also expect to see fidelity measures. We applaud these trends.

The use of fidelity measures in practice settings can have salutary effects on quality improvement. One litmus test that a fidelity measure has become widely accepted occurs when individual fidelity checklist items become the basis for policy discussions around what the standards should be. Insofar as these discussions become centered on specific data and empirical criteria, these debates represent a constructive advance over policy driven by politics.

In this paper we have sought to illustrate the advantages of using fidelity measures and the disadvantages of not using them. Clearly, the demand for these measures has grown in response to the problems that have emerged when fidelity has been ignored. Program implementation, of course, is not the only important element in facilitating effective psychiatric rehabilitation. Many other factors must be considered, including adequate resource allocation, staff competencies, and specification of target populations. Adequate funding is a necessary, but clearly not sufficient condition to ensure adequate services. Staff competencies are obviously also critically important to a program's success. No program, if staffed by inexperienced or demoralized workers, can be effective, yet this obvious fact too often is overlooked or ignored in practice. Specification of the target population is still another neglected aspect of program practice. Practice guidelines, including the PORT recommendations, have sought to define phase of illness and characteristics of clients for whom interventions are effective. To date, fidelity scales generally have not incorporated items relating to client characteristics.

We have made the leap in this paper of linking research and practical uses of fidelity measures. Some readers will argue that we have blurred some distinctions between different types of monitoring activities. Certainly there is a continuum of monitoring activities ranging from the highly detailed program monitoring and fidelity measurement that should be done in a randomized controlled trial to brief surveys that may be suitable for monitoring implementation of a new statewide program model. We propose that there are advantages to thinking of these activities as more closely related than they have been historically viewed.

The most significant influence of the growing emphasis on fidelity measurement is the pressure to define program models operationally. Historically, psychiatric rehabilitation has been a field that has prided itself in its innovation, creativity, and flexibility. Not everyone is keen on the idea of program models. IAPSRS, for example, has pointed out an alarming trend for state and local mental health authorities to fall prey to the "single model trap" (Hughes & Clement, 1999; IAPSRS, 1997a). The implicit assumption with an emphasis on a single model is that "one shoe fits all." When policymakers stipulate that funding be provided only for a specific model of services, they foreclose other options. In response, providers often argue that their existing eclectic programs are equally or more effective for *their* clients, or that the designated model is inappropriate for some segments of their population. According to this reasoning, the idea of a single program model is antithetical to the idea of a "flexible array of options" that has been at the heart of the psychiatric rehabilitation philosophy.

Another historical objection to promulgating program models is given by Bachrach (1988), who argued that model programs are developed in a par-

ticular sociocultural and economic context that do not generalize to local conditions. Providers know their constituencies best, and they should adapt the models to their conditions, rather than mechanically apply what the original developers did. This particular argument has been applied repeatedly to the Madison model of ACT (Stein & Test, 1980) over the past 20 years.

We agree that these ''anti-model'' viewpoints have merit. Unfortunately, such arguments can be conveniently used to maintain the status quo, to reassure providers that whatever they are doing is adequate. Sometimes programs follow no apparent model at all, with individual clinicians varying widely in their intervention strategies. Another common response by providers is to blend an existing approach with the new model, accepting elements of the new model while discarding others. If the philosophy of operationally defining program models and then testing their effectiveness is applied to alternative approaches, then we see no conflict between the ''model'' viewpoint and the viewpoint of critics.

Most fidelity measures in the psychiatric rehabilitation field are rudimentary. We have our work cut out for us if we want to pursue the agenda suggested above. We believe that we should apply the lessons from the psychotherapy fidelity literature, as well as the broader literature on measurement. Although fidelity measurement is no panacea, it can help in both the research and practice of psychiatric rehabilitation.

ACKNOWLEDGMENT

Work on this paper was supported by a grant from the Human Services Research Institute and Grant MH00842 from the National Institute of Mental Health.

REFERENCES

Allness, D. J., & Knoedler, W. H. (1998). *The PACT model of community-based treatment for persons with severe and persistent mental illness: A manual for PACT start-up.* Arlington, VA: NAMI.

Anthony, W. A. (1994). Whither the ''Boston University model''? *Psychosocial Rehabilitation Journal, 17*(4), 169–170.

Anthony, W. A., Cohen, M. R., & Farkas, M. D. (1982). A psychiatric rehabilitation program: Can I recognize one when I see one? *Community Mental Health Journal, 18,* 83–96.

Anthony, W. A., Cohen, M. R., & Pierce, R. M. (1980). *Instructor's guide to the psychiatric rehabilitation practice series.* Baltimore, MD: University Park Press.

Bachrach, L. L. (1988). The chronic patient: On exporting and importing model programs. *Hospital and Community Psychiatry, 39,* 1257–1258.

Barton, R. (1997). *Psychosocial rehabilitation services and community support systems: Outcomes research and policy implications.* Springfield, IL: Office of Mental Health of the Illinois Department of Human Services.

Beard, J. H., Propst, R. N., & Malamud, T. J. (1982). The Fountain House model of rehabilitation. *Psychosocial Rehabilitation Journal, 5*(1), 47–53.

Becker, D. R., & Drake, R. E. (1993). *A working life: The Individual Placement and Support (IPS) Program.* Concord, NH: New Hampshire-Dartmouth Psychiatric Research Center.

Becker, T., Holloway, F., McCrone, P., & Thornicroft, G. (1998). Evolving service interventions in Nunhead and Norwood: PRiSM Psychosis Study 2. *British Journal of Psychiatry, 173,* 371–375.

Bond, G. R. (1991). Variations in an assertive outreach model. *New Directions for Mental Health Services, 52,* 65–80.

Bond, G. R., Becker, D. R., Drake, R. E., & Vogler, K. M. (1997). A fidelity scale for the Individual Placement and Support model of supported employment. *Rehabilitation Counseling Bulletin, 40,* 265–284.

Bond, G. R., Drake, R. E., Becker, D. R., & Mueser, K. T. (1999a). Effectiveness of psychiatric rehabilitation approaches for employment of people with severe mental illness. *Journal of Disability Policy Studies, 10*(1), 18–52.

Bond, G. R., Evans, L., Kim, H., & Goodman, C. (1999b). *Evaluation of Connect98. Phase II: Site visits by network staff.* Final report to Illinois Office of Mental Health. Indianapolis, IN: Indiana University-Purdue University Indianapolis.

Bond, G. R., Miller, L. D., Krumwied, R. D., & Ward, R. S. (1988). Assertive case management in three CMHCs: A controlled study. *Hospital and Community Psychiatry, 39,* 411–418.

Bond, G. R., Picone, J., Mauer, B., Fishbein, S., & Stout, R. (1999). The Quality of Supported Employment Implementation Scale. In G. Revell, K. J. Inge, D. Mank, & P. Wehman (Eds.), *The impact of supported employment for people with significant disabilities: Preliminary findings from the National Supported Employment Consortium* (pp. 73–88). Richmond, VA: Virginia Commonwealth University.

Brekke, J. S. (1987). The model-guided method for monitoring program implementation. *Evaluation Review, 11*(3), 281–299.

Brekke, J. S. (1988). What do we really know about community support programs? Strategies for better monitoring. *Hospital and Community Psychiatry, 39,* 946–952.

Brekke, J. S., & Aisley, R. A. (1990). The client interaction scale. *Evaluation and the Health Professions, 13,* 215–226.

Brekke, J. S., & Test, M. A. (1987). An empirical analysis of services delivered in a model community support program. *Psychosocial Rehabilitation Journal, 10*(4), 51–61.

Brekke, J. S., & Test, M. A. (1992). A model for measuring the implementation of community support programs: Results from three sites. *Community Mental Health Journal, 28,* 227–247.

Brekke, J. S., & Wolkon, G. H. (1988). Monitoring program implementation in community mental health settings. *Evaluation and the Health Professions, 11,* 425–440.

Burns, T., Creed, F., Fahy, T., Thompson, S., Tyrer, P., & White, I. (1999). Intensive versus standard case management for severe psychotic illness: A randomised trial. UK 700 Group. *Lancet, 353,* 2185–2189.

Burt, M., Duke, A. E., & Hargreaves, W. A. (1998). The Program Environment Scale: Assessing client perceptions of community-based programs for the severely mentally ill. *American Journal of Community Psychology, 26,* 853–879.

Cochrane, J., Durbin, J., & Goering, P. (1997). *Best practices in mental health reform: Discussion paper.* Toronto, ON: Clarke Institute of Psychiatry.

Deci, P. A., Santos, A. B., Hiott, D. W., Schoenwald, S., & Dias, J. K. (1995). Dissemination of assertive community treatment teams. *Psychiatric Services, 46,* 676–678.

Dewan, N. A., & Carpenter, D. (1997). Performance measurement in healthcare delivery systems. In L. J. Dickstein, M. B. Riba, & J. M. Oldham (Eds.), *American Psychiatric Press Review of Psychiatry* (Vol. 16, pp. 81–102). Washington, DC: American Psychiatric Press.

Dincin, J. (1995). A pragmatic approach to psychiatric rehabilitation: Lessons from Chicago's Thresholds program. *New Directions for Mental Health Services, 69*(Whole Issue).

Drake, R. E., McHugo, G. J., Becker, D. R., Anthony, W. A., & Clark, R. E. (1996). The New Hampshire study of supported employment for people with severe mental illness: Vocational outcomes. *Journal of Consulting and Clinical Psychology, 64,* 391–399.

Eysenck, H. (1952). The effects of psychotherapy, an evaluation. *Journal of Consulting Psychology, 16,* 319–324.

Fishbein, S. M. (1988). Partial care as a vehicle for rehabilitation of individuals with severe psychiatric disability. *Rehabilitation Psychology, 33,* 57–64.

Giesler, L. J., & Hodge, M. (1998). Case management in behavioral health care. *International Journal of Mental Health, 27,* 26–40.

Gold Award. (1999). The wellspring of the clubhouse model for social and vocational adjustment of persons with serious mental illness: Fountain House, New York City. *Psychiatric Services, 50,* 1473–1476.

Gorey, K. M., Leslie, D. R., Morris, T., Carruthers, W. V., John, L., & Chacko, J. (1998). Effectiveness of case management with severely and persistently mentally ill people. *Community Mental Health Journal, 34,* 241–250.

Hall, L. L., Edgar, E. R., & Flynn, L. M. (1997). *Stand and deliver: Action call to a failing industry.* Arlington, VA: National Alliance for the Mentally Ill.

Hargreaves, W. A., Shumway, M., Hu, T. W., & Cuffel, B. (1998). *Cost-outcome methods for mental health.* San Diego, CA: Academic Press.

Heflinger, C. A. (1996). Implementing a system of care: Finding from the Fort Bragg Evaluation Project. *Journal of Mental Health Administration, 23,* 16–29.

Henggeler, S. W., Pickrel, S. G., & Brondino, M. J. (1999). Multisystemic treatment of substance-abusing and -dependent delinquents: Outcomes, treatment fidelity, and transportability. *Mental Health Services Research, 1,* 171–184.

Herz, M. I., Liberman, R. P., Lieberman, J. A., Marder, S. R., McGlashan, T. H., Wyatt, R. J., & Wang, P. (1997). Practice guideline for the treatment of patients with schizophrenia. *American Journal of Psychiatry, 154*(Supplement), 1–63.

Hogan, M. F. (1999). Public-sector mental health care: New challenges. *Health Affairs, 18*(5), 106–111.

Hogarty, G. E. (1995). Schizophrenia and modern mental health services. *Decade of the Brain, 6*(1), 3–6.

Hughes, R. (1999). The meaning of ''evidence based'' services in PSR. *IAPSRS Connection, 2*(1) 10–12.

Hughes, R., & Clement, J. (1999, March 1). Time to end the model wars. *IAPSRS Connection, 1,* 1.

IAPSRS. (1997a). *IAPSRS position paper: The single model trap.* Columbia, MD: International Association of Psychosocial Rehabilitation Services.

IAPSRS. (1997b). *Practice guidelines for the psychiatric rehabilitation of persons with severe and persistent mental illness in a managed care environment.* Columbia, MD: International Association of Psychosocial Rehabilitation Services.

Jerrell, J. M., & Hargreaves, W. A. (1991). *The operating philosophy of community programs* (Working Paper 18). Berkeley, CA: Institute for Mental Health Services Research.

Jerrell, J. M., & Ridgely, M. S. (1999). Impact of robustness of program implementation on outcomes of clients in dual diagnosis programs. *Psychiatric Services, 50,* 109–112.

Johnsen, M., Samberg, L., Calsyn, R., Blasinsky, M., Landow, W., & Goldman, H. (1999). Case management models for persons who are homeless and mentally ill: The ACCESS Demonstration Project. *Community Mental Health Journal, 35,* 325–346.

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

Latimer, E. (1999a). *Conseil d'évaluation des technologies de la santé du Québec. Suivi intensif en équipe dans la communauté pour personnes atteintes de troubles mentaux graves* (CETS 99-1 RF). Montreal: CETS.

Latimer, E. (1999b). Economic impacts of assertive community treatment: A review of the literature. *Canadian Journal of Psychiatry, 44,* 443–454.

Lehman, A. F., Steinwachs, D. M., & PORT Co-Investigators. (1998). At issue: Translating research into practice: The Schizophrenia Patient Outcomes Research Team (PORT) treatment recommendations. *Schizophrenia Bulletin, 24,* 1–10.

Leithwood, K. A., & Montgomery, D. J. (1980). Evaluating program implementation. *Evaluation Review, 4*(2), 193–214.

Lieberman, M. A., Yalom, I. D., & Miles, M. B. (1973). *Encounter groups: First facts.* New York: Basic Books.

Lipsey, M. W. (1990). *Design sensitivity.* Newbury Park, CA: Sage.

Marshall, M., & Creed, F. (in press). Assertive community treatment: Is it the future of community care in the UK? *Social Psychiatry and Psychiatric Epidemiology.*

Marshall, M., Lockwood, A., Green, R., & Gray, A. (1998). Case management for people with severe mental disorders (Cochrane Review), *Cochrane Library* (Updated quarterly ed.). Oxford, England: Update Software.

McCarthy, D., Thompson, D., & Olson, S. (1998). Planning a statewide project to convert day treatment to supported employment. *Psychiatric Rehabilitation Journal, 22*(1), 30–33.

McEvoy, J. P., Scheifler, P. L., & Frances, A. (1999). The expert consensus guideline series: Treatment of schizophrenia 1999. *Journal of Clinical Psychiatry, 11*(Supplement), 1–80.

McGrew, J. H., & Bond, G. R. (1995). Critical ingredients of assertive community treatment: Judgments of the experts. *Journal of Mental Health Administration, 22,* 113–125.

McGrew, J. H., Bond, G. R., Dietzen, L. L., & Salyers, M. P. (1994). Measuring the fidelity of implementation of a mental health program model. *Journal of Consulting and Clinical Psychology, 62,* 670–678.

McHugo, G. J., Drake, R. E., Teague, G. B., & Xie, H. (1999). The relationship between model fidelity and client outcomes in the New Hampshire Dual Disorders Study. *Psychiatric Services, 50,* 818–824.

McHugo, G. J., Hargreaves, W. A., Drake, R. E., Clark, R. E., Xie, H., Bond, G. R., & Burns, B. J. (1998). Methodological issues in assertive community treatment studies. *American Journal of Orthopsychiatry, 68,* 246–260.

Mitchell, K. M., Bozarth, J. D., & Krauft, C. C. (1977). A reappraisal of the therapeutic effectiveness of accurate empathy, nonpossessive warmth, and genuineness. In A. S. Gurman & A. M. Razin (Eds.), *Effective psychotherapy: A handbook of research* (pp. 482–502). Oxford: Pergamon.

Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review, 11,* 247–266.

Moos, R. H. (1974). *Community Oriented Program Environment Scale.* Palo Alto, CA: Consulting Psychologist Press.

Mowbray, C. T., Plum, T. B., & Masterton, T. (1998). Harbinger II: Deployment and evolution of assertive community treatment in Michigan. *Administration and Policy in Mental Health, 25,* 125–139.

Moxley, D. P. (1993). Clubhouse standards as a program development tool. *Psychsocial Rehabilitation Journal, 17*(2), 177–183.

Mueser, K. T., Bond, G. R., Drake, R. E., & Resnick, S. G. (1998). Models of community care for severe mental illness: A review

of research on case management. *Schizophrenia Bulletin, 24,* 37–74.

Mueser, K. T., & Glynn, S. M. (1999). *Behavioral family therapy for psychiatric disorders.* (2nd ed.). Oakland, CA: New Harbinger.

Noble, J. H. (1991). *The benefits and costs of supported employment for people with mental illness and with traumatic brain injury in New York State* (C-0023180). Buffalo, NY: Research Foundation of the State University of New York.

Orlinsky, D. E., Grawe, K., & Parks, B. K. (1994). Process and outcome in psychotherapy—Noch einmal. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 270–376). New York: Wiley.

Paul, G. L., & Lentz, R. J. (1977). *Psychosocial treatment of chronic mental patients: Milieu versus social learning programs.* Cambridge, MA: Harvard University Press.

Propst, R. (1992). Standards for clubhouse programs: Why and how they were developed. *Psychosocial Rehabilitation Journal, 16*(2), 25–30.

Rapp, C. A. (1999). *Best Practice Fidelity Tools.* Lawrence, KS: University of Kansas School of Social Welfare.

Rezmovic, E. L. (1984). Assessing treatment implementation amid the slings and arrows of reality. *Evaluation Review, 8,* 187–204.

Rogers, C. R. (1951). *Client-centered therapy.* Boston: Houghton Mifflin.

Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology, 21,* 95–103.

Rosenheck, R., Neale, M., Leaf, P., Milstein, R., & Frisman, L. (1995). Multisite experimental cost study of intensive psychiatric community care. *Schizophrenia Bulletin, 21,* 129–140.

Salyers, M. P., Masterton, T. W., Fekete, D. M., Picone, J. J., & Bond, G. R. (1998). Transferring clients from intensive case management: Impact on client functioning. *American Journal of Orthopsychiatry, 68,* 233–245.

Schaedle, R., & Epstein, I. (2000). Specifying intensive case management: A multiple perspective approach. *Mental Health Services Research,* this issue.

Schreier, M. A., & Rezmovic, E. L. (1983). Measuring the degree of program implementation: A methodological review. *Evaluation Review, 7,* 599–633.

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

Solomon, P. (1992). The efficacy of case management services for severely mentally disabled clients. *Community Mental Health Journal, 28,* 163–180.

Stein, L. I., & Santos, A. B. (1998). *Assertive community treatment of persons with severe mental illness.* New York: Norton.

Stein, L. I., & Test, M. A. (1980). An alternative to mental health treatment. I: Conceptual model, treatment program, and clinical evaluation. *Archives of General Psychiatry, 37,* 392–397.

Suh, C. S., O'Malley, S. S., & Strupp, H. H. (1986). The Vanderbilt process measures: The Psychotherapy Process Scale (VPPS) and the Negative Indicators Scale (VNIS). In L. S. Greenberg & W. M. Pinsof (Eds.), *The psychotherapeutic process: A research handbook* (pp. 285–323). New York: Guilford Press.

Teague, G. B., Bond, G. R., & Drake, R. E. (1998). Program fidelity in assertive community treatment: Development and use of a measure. *American Journal of Orthopsychiatry, 68,* 216–232.

Teague, G. B., Drake, R. E., & Ackerson, T. H. (1995). Evaluating use of continuous treatment teams for persons with mental illness and substance abuse. *Psychiatric Services, 46,* 689–695.

Torrey, E. F., Erdman, K., Wolfe, S. M., & Flynn, L. M. (1990). *Care of the seriously mentally ill: A rating of state programs* (3rd ed.). Arlington, VA: National Alliance for the Mentally Ill.

Torrey, W. C., & Wyzik, P. F. (1997). *New Hampshire clinical practice guidelines for adults in community support programs.* Lebanon, NH: West Central Services.

Turner, J. C., & TenHoor, W. J. (1978). The NIMH community support program: Pilot approach to a needed social reform. *Schizophrenia Bulletin, 4,* 319–348.

Wallace, C. J., Liberman, R. P., MacKain, S. J., Blackwell, G., & Eckman, T. A. (1992). Effectiveness and replicability of modules for teaching social and instrumental skills to the severely mentally ill. *American Journal of Psychiatry, 149,* 654–658.

Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology, 61,* 620–630.

Wang, Q., Macias, C., & Jackson, R. (1999). First step in the development of a clubhosue fidelity instrument: Content analysis of clubhouse certification reports. *Psychiatric Rehabilitation Journal, 22*(3), 294–301.

Zahrt, D. M., Bond, G. R., Salyers, M. P., & Teague, G. B. (1999). *Dartmouth ACT Fidelity Scale: A practical application in a statewide project.* Unpublished paper. Indiana University-Purdue University Indianapolis, Indianapolis, Indiana.