PROGRAM FIDELITY
IN ASSERTIVE COMMUNITY TREATMENT:
Development and Use of a Measure

Gregory B. Teague, Ph.D., Gary R. Bond, Ph.D., Robert E. Drake, M.D., Ph.D.

Assertive community treatment (ACT) is a complex community-based service approach to helping people with severe mental disorders live successfully in the community. Effective replication of the model and research on critical elements require explicit criteria and measurement. A measure of program fidelity to ACT and the results of its application to fifty diverse programs are presented.

Assertive community treatment (ACT) has long been recommended as an effective treatment for people with severe mental disorders (Drake & Burns, 1995). Since its original development in Madison, Wisconsin as Training in Community Living, a community-based alternative to psychiatric hospital care (Stein & Test, 1980), the program has been widely emulated and frequently evaluated (Burns & Santos, 1995; Olfson, 1990; Thompson, Griffith, & Leaf, 1990). The general model has been adapted and evaluated in various settings and with various subgroups of people with severe mental disorders, including young adults with recent-onset schizophrenia (Test, 1992), people with co-occurring alcohol and drug disorders (Drake, Teague, & Warren, 1990; Teague, Drake, & Ackerson, 1995), veterans (Rosenheck, Neale, Leaf, Milstein, & Frisman, 1995), and people who are homeless (Dixon, Krauss, & Kernan, 1995; Morse, Calsyn, & Allen, 1992), as well as with families (McFarlane, Stasny, & Deakin, 1992).

Several publications have described the rationale, important features, and operating principles of ACT (Santos, 1993; Test, 1992; Test & Stein, 1976). However, despite longstanding acclaim and demonstrated effectiveness of the model, only recently have efforts been made to define operationally and measure empirically its critical dimensions (Allness and Knoedler, 1998; McGrew & Bond, 1995; McGrew, Bond, & Dietzen, 1994; Teague et al., 1995). In the absence of explicit model criteria, explanations for variation in outcome must remain speculative. For proper interpretation of outcome findings, an adequate measure of fidelity to the ACT model is needed. This article describes the development and results of using the Dartmouth ACT Scale.
TEAGUE ET AL

(DACTS), a product of recent efforts to make the essential features of ACT operationally explicit.*

MEASUREMENT OF FIDELITY

A number of authors have decried the lack of consistency in providing careful program documentation with reports of the effectiveness of interventions, including treatment (Moncher & Prinz, 1991) and service models (Brekke, 1988; Brekke & Test, 1992). Without detailed descriptions of interventions, replication is difficult; without reliable measurement of interventions, conclusions about presence or absence of effects are questionable. Careful specification of the critical components of a model, and use of measures based on operational definitions of these components provide a means both to inhibit drift away from the model and to evaluate the respective contributions of theoretically distinct components (Bickman, 1987, 1990; Bond, 1991; Chen, 1990; Finney & Moos, 1989; Lipsey, 1990). Program measurement using multiple variables and scales also permits empirical decomposition of programs as a supplement to theory-driven analyses (Scott & Sechrest, 1989). Empirical measures of program implementation can be used in several ways: to shape program performance in formative stages; to serve as measures of dosage or treatment strength; or to define inclusion criteria for multisite studies (Teague et al., 1995).

Several approaches have been used to develop quantitative measures of program implementation or fidelity of services for people with severe mental illness. Brekke (1987) reported use of staff time-logs incorporating categories of activity that were specifically matched to critical aspects of the intervention. Jerrell and Hargreaves (1991) developed the Community Program Philosophy Scale to measure a wide range of community support programs. This instrument uses staff ratings of a program’s values and practices on 20 four-item scales reflecting relevant dimensions of community support philosophy. Brekke and Test (1987) used a battery of existing instruments for empirical discrimination of programs.

More recently, and more specifically related to ACT, McGrew et al. (1994) surveyed experts to identify critical ingredients of this model. Using the identified ingredients, they then developed a program survey to elicit data on a selected set of program characteristics and performance. Fidelity variables were grouped in three a priori dimensions: staffing, organization, and service. Program and client hospitalization data were collected from a number of ACT programs in the study. McGrew and colleagues were then able to demonstrate that programs with greater fidelity to the ACT model, as defined in terms of the subset of identified critical ingredients, were also more effective in reducing hospital use.

Teague et al. (1995) drew from the literature on ACT (Bond, 1991; Santos, 1993; Test, 1992; Test & Stein, 1976) to identify 13 key dimensions of program implementation in the study of an ACT-like intervention for people with severe mental disorders and co-occurring substance use disorders. Dimensions reflected features both of the general ACT model and of a particular model of integrated treatment for the two sets of disorders (Drake et al., 1990). The authors defined one or more indicators for each dimension to map onto an anchored scale, using multiple methods and data sources. Data on outcomes were not then available, but it was evident that differences in program fidelity across experimental and control conditions were most pronounced for those programmatic features that were the most structurally constrained, e.g., caseload size and composition. Greater model drift had occurred for more discretionary features, such as overall treatment approach and in vivo services.

*Copies of the full scale and instructions are available from the first author.
The DACTS, reported here extends the general methods of this approach (Teague et al., 1995).

Construction of Dartmouth ACT Scale

Twenty-six program criteria for fidelity to ACT were derived from multiple sources, including literature describing the model (Santos, 1993; Test, 1992; Test & Stein, 1976), results from previous work on fidelity (McGrew et al., 1994; Teague et al., 1995), and expert opinion (McGrew & Bond, 1995). The majority of criteria and anchors were adapted from variables used as indicators for the 13 dimensions reported in Teague et al. (1995). Additional variables were designed on the basis of results reported in McGrew et al. (1994). As in McGrew et al., criteria were grouped for convenience in three a priori dimensions, although these dimensions would not necessarily be reflected in the empirical structure. Colleagues studying ACT or related programs provided helpful suggestions for refinement of items (T. Ackerson, D. Allness, W. Breakey, B. Burns, R. Calsyn, P. Deci, L. Dixon, S. Essock, N. Kontos, A. Lehman, M. Neale, R. Rosenheck, & A. Santos, personal communication, July-August, 1995). Later editorial refinements for the version shown in this article were suggested in discussion with site evaluators from the ACCESS program (Randolph, 1995).

Table 1 identifies and lists descriptors for the 26 program criterion variables grouped in the three dimensions. (Two variables, H11 and S10, added after the study, are also included for a total of 28 criteria.) Overall, the current version of the DACTS focuses primarily on structural aspects of the ACT model, for which behav-

<table>
<thead>
<tr>
<th>HUMAN RESOURCES STRUCTURE/COMPOSITION</th>
<th>05 Responsibility for Hospital Admissions program is involved in hospital admissions</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1 Small Caseload client/provider ratio of 10:1</td>
<td>06 Responsibility for Hospital Discharge Planning program is involved in planning hospital discharges</td>
</tr>
<tr>
<td>H2 Team Approach provider group functions as team rather than individual practitioners; clinicians know &amp; work with all clients</td>
<td>07 Time-Unlimited Services program closes no cases, remains point-of-contact for all clients as needed</td>
</tr>
<tr>
<td>H3 Program Meeting program meets frequently to plan, review services for each client</td>
<td>NATURE OF SERVICES</td>
</tr>
<tr>
<td>H4 Practicing Team Leader: supervisor of front-line clinicians provides direct services.</td>
<td></td>
</tr>
<tr>
<td>H5 Continuity of Staffing program maintains same staffing over time.</td>
<td>S1 In-Vivo Services program monitors status, develops community living skills in community rather than office</td>
</tr>
<tr>
<td>H6 Staff Capacity: program operates at full staffing.</td>
<td>S2 No Dropout Policy program engages/retains clients at mutually satisfactory level</td>
</tr>
<tr>
<td>H7 Psychiatrist on Staff at least one full-time psychiatrist per 100 clients assigned to program</td>
<td>S3 Assertive Engagement Mechanisms uses street outreach, plus legal mechanisms (e.g., representative payees, probation/parole, OP commitment) as indicated.</td>
</tr>
<tr>
<td>H8 Nurse on Staff ≥2 full-time nurses per 100 clients.</td>
<td>S4 Intensity of Service high total amount of service time, as needed.</td>
</tr>
<tr>
<td>H9 Substance Abuse Specialist on Staff ≥2 staff with 1 yr training/clinical exp. in substance abuse treatment</td>
<td>S5 Frequency of Contact high number of service contacts, as needed.</td>
</tr>
<tr>
<td>H10 Vocational Specialist on Staff ≥1 staff member with ≥1 yr training/exp in vocational rehab/support.</td>
<td>S6 Work With Support System with or without client present, program provides support/skills for client’s support network: family, landlords, employers</td>
</tr>
<tr>
<td>H11 Program Size: sufficient absolute size to provide consistently the necessary staff diversity &amp; coverage (Data on this variable not collected in current study)</td>
<td>S7 Individualized Substance Abuse Treatment ≥1 program member provides direct treatment &amp; substance abuse treatment for clients w/substance use disorders</td>
</tr>
<tr>
<td>ORGANIZATIONAL BOUNDARIES</td>
<td>S8 Dual Disorder Treatment Groups group modalities used as tx strategy for people w/substance disorders</td>
</tr>
<tr>
<td>O1 Explicit Admission Criteria: clearly identified mission to serve particular population, measurable, operationally defined criteria to screen out inappropriate referrals.</td>
<td>S9 Dual Disorders Model: uses a stage-wise treatment model that is nonconfrontational, follows behavioral principles, considers interactions of mental illness &amp; substance abuse, &amp; has gradual expectations of abstinence</td>
</tr>
<tr>
<td>O2 Intake Rate: takes clients in at a low rate to maintain a stable service environment</td>
<td>S10 Role of Consumers on Treatment Team: clients involved as team members providing direct services. (Data on this variable not collected in current study.)</td>
</tr>
<tr>
<td>O3 Full Responsibility for Treatment Services, as well as case management/psychiatric services, program directly provides counseling/psychotherapy, housing support, substance abuse, employment, &amp; rehab services</td>
<td></td>
</tr>
<tr>
<td>O4 Responsibility for Crisis Services 24-hour coverage of psychiatric crises</td>
<td></td>
</tr>
</tbody>
</table>

TABLE 1

PROGRAM CRITERIA FOR FIDELITY TO ACT
ioral anchors may be more readily developed, rather than on more clinical aspects that may be equally important to effectiveness but are harder to measure. For example, although individualization of treatment is an important feature of ACT, it is quite difficult to operationalize adequately, and no such item is currently included.

Structure and content. As can be seen in Table 1, the first dimension, human resources, addresses composition and structure of program staffing: ratio of clients to staff; team functioning, including caseload sharing, regular meetings, and leadership by a practicing clinician; and specific disciplines or specialized staffing, including physician, nurse, and substance abuse and vocational specialists. Although full and continuous staffing are not necessarily limited to the ACT model, they are specified because failure to achieve them seriously compromises program functioning.

The second dimension, organizational boundaries, addresses program responsibility and relationships with other programmatic components. Criteria include specificity of admission criteria; rate at which the program takes on new clients; responsibility for specific types of rehabilitative and crisis services; movement in and out of hospitals; and commitment to unlimited services, as needed.

The third dimension, nature of services, addresses the range and nature of services and overall treatment approach. Criteria specify services in the community; outreach and follow-up to maintain clients in service; explicit use of assertive engagement techniques; intensive and frequent services, as needed; staff working with clients' support systems; and treatment for substance abuse problems as needed, on an individualized basis and through a non-confrontational, stage-wise approach using group modalities.

Item anchors and scoring. Two general steps were followed in operationalizing each criterion. First, an indicator was identified and defined in terms of the nature and source of data. In order to satisfy a given criterion, empirical evidence from a particular source would need a certain value. Selection of indicators was guided, in part, by feasibility of definition and measurement. When use of quantitative indicators to represent a particular criterion would typically be infeasible, qualitative indicators were specified. Most criteria, such as frequency of team meetings, have indicators that fully represent the construct. A few criteria were operationalized less completely. For example, provision of in vivo services is measured simply by the proportion of time that staff spend providing services out of the office. While this distinction does not fully capture the construct of spending time with clients in the places most relevant for them, it defines a pertinent and easily measurable aspect of that construct.

The second operationalizing step entailed specifying anchors for each point on the rating scale. Five points were selected to provide an efficient number of discrete points that would still approximate a continuous scale. The most critical points for definition were the high and low ones. A central function of a fidelity measure like this is to distinguish programs conforming to the ideal from those conforming to norms for the corresponding conventional treatment. Accordingly, because the Madison PACT team has long been considered the paradigm of ACT, high anchors were derived as much as possible from published norms for this program (Brekke & Test, 1987). Others were based on descriptors in published program literature (Test, 1992; Santos, 1993), on a survey of ACT experts (McGrew & Bond, 1995), and on author consensus. Because the context for ACT programs—and for alternative treatment for clients—is typically services-as-usual, not the total absence of services, low anchors for quantitative variables were set to an estimated minimal value of typical case-management-based services, rather than to an extreme value, such as zero.
Thus, the low anchor for annual caseload retention is defined as 50%. Similarly, scores on the item for small caseload can range from a high of five for a client:staff ratio of ten or less, to a low of one for any ratio higher than 80. Intermediate points were interpolated and adjusted to obtain distribution across the scale. Overall, this strategy was intended to ensure full use of the scale and high variance in cross-model measurement.

Programs ratings are generated by mapping program data or features onto anchors for each item. Table 2 illustrates anchors for three items, one from each a priori dimension. For quantitative items, program data or results of specified calculations map directly onto ranges specified in anchors. On item S1 in Table 2, for example, a program providing 50% of its direct services in the community would be rated at 3. For qualitative items, program features are compared with anchors, and the value corresponding to the best fit is assigned. On item O4 in Table 2, for example, a program that provides only consultation via telephone for emergency services would be rated at 3.

Data sources. The DACTS is designed for use by informed raters. Information to be used in making ratings ideally comes from a range of sources, as in the antecedent measures (McGrew, et al., 1994; Teague et al., 1995). These sources include reports of program behavior from program supervisors or staff; documents reflecting program authority, responsibility, policies, and procedures; management information system or other sources of quantitative data on staffing, clientele, and services. Instructions for using the scale include recommendations regarding the different sources that are relevant for different items. When the primary rater lacks direct access to all the data required, a composite rating can be derived from structured interviews with multiple informants who have direct access to some of the data. This approach was adopted in one of the sets of studies reported here, where raters were initially less informed about program details than were raters in other studies. Informant categories were specified in order to provide the necessary coverage of content (R. Calsyn, H. Goldman; personal communication, November 1995).

Table 2

<table>
<thead>
<tr>
<th>ITEM</th>
<th>SCALE</th>
</tr>
</thead>
<tbody>
<tr>
<td>H3. Program Meeting. Program meets frequently to plan &amp; review services for each client</td>
<td>1</td>
</tr>
<tr>
<td>Service planning for each client usually occurs once a month or less frequently</td>
<td>At least twice a month but less often than once a week</td>
</tr>
<tr>
<td>O4. Responsibility for Crisis Services. Program has 24-hr responsibility for psychiatric crises</td>
<td>No responsibility for handling crises after hours</td>
</tr>
<tr>
<td>S1. In-Vivo Services. Program works to monitor status and develop living skills in community rather than in office</td>
<td>Less than 20% time in community</td>
</tr>
</tbody>
</table>
TEAGUE ET AL

METHOD

Study Design

The research design consisted of examining four groups of programs implemented with varying degrees of intended fidelity to the ACT model. One group was intended to be in close replication of the ACT model; a second group included intensive case management programs based on ACT principles but with some modified elements; a third group incorporated selected features of ACT for a specific population; and a fourth group represented more traditional case management services. The four groups were examined together, employing the "method of known groups" (Cronbach & Meehl, 1955; Pedhazur & Schmelkin, 1991) as an approach to establishing the construct validity of a new scale. This method tests whether an assessment approach yields differences corresponding to a pre-existing classification. In this case, it was expected that the first group would have higher scores on the fidelity scale than the programs more loosely based on the ACT model, and that these, in turn, would have higher scores than the traditional case management programs.

Program Models

A total of 50 programs in four groups of models were evaluated using the DACTS. All programs were designed to serve persons with severe mental illness. The first group of programs consisted of 14 different sites within five separate studies of effectiveness of ACT (Dixon, et al., 1995; Drake et al., 1998; Essock & Kontos, 1995; Primm, 1996; Santos, Deci, Lachance, 1993). Two studies included one site each; two included two sites each; and one included seven sites. All were designed to implement as nearly as possible the essential features of the ACT model and were closely supervised by researchers, albeit with varying degrees of formal authority over program implementation. The study incorporating seven sites was designed to evaluate the addition of integrated treatment for co-occurring substance use disorders to the basic ACT model (Drake, et al., 1998, this issue; Teague, et al., 1995). The studies in this group began during 1989–1990, had follow-up periods ranging from 18 months to three years, and conducted final follow-up interviews for the cohorts during 1992–1995. Program size ranged from 40 to 75 clients, and caseload ratios ranged from seven to 13 clients per clinician.

The second group of programs consisted of ten sites from the VA Intensive Psychiatric Community Care (IPCC) program, an ACT-like intensive case management program implemented in the Department of Veteran Affairs for persons with severe mental illness who had been high users of hospital services (Rosenheck & Neale, in press; Rosenheck, Neale, Leaf, Milstein, & Frisman, 1995). Six of the ten sites were general medical and surgical hospitals, four were neuropsychiatric hospitals; all were implemented and evaluated during the period 1987–1995. The model was similar to the Madison PACT model in several respects: it was explicitly designed to provide the clinically indicated degree of service intensity for each patient; program size was approximately 45, and caseload ratios ranged from 7:1 to 15:1; services were provided as much as possible in community settings; emphasis was placed on practical problem solving and skill building through a broad range of rehabilitative services; and continuity of care was assured by assertive maintenance of contact with and responsibility for each patient. IPCC was different from PACT in its lesser emphasis on a team focus, greater staff dependence on other programs, and greater patient involvement in other service programs.

The third group consisted of 15 of the 18 sites in the Access to Community Care and Effective Services and Supports (ACCESS) program, sponsored by the Center for Mental Health Services (Randolph, 1995; Rosenheck et al., in press). The goal of ACCESS is to evaluate the impact of increases in system integration on homeless persons
with severe mental illness. The nine pairs of sites were funded in the fall of 1993, and service programs were established during the first year. Individual service programs were not intended to be replications of ACT, but they shared many of its features, including multidisciplinary team organization, assertive treatment approaches, and extensive service responsibility. Because of these similarities, evaluators of this national-level intervention adopted the DACTS as a measure to compare implementation across sites. Intentional departures from the usual ACT model included an explicit expectation of time-limited client length-of-stay and, in many cases, structural separation of outreach functions. Each site was expected to engage and serve a new cohort of 100 clients every year.

The fourth group consisted of 11 of the programs that served as control sites for the ACT studies in the first group. This group used standard case management models in comprehensive community mental health treatment settings, and had been part of ongoing service programs for many years; caseload ratios ranged from 25 to 60 clients per clinician. Control group data from the two studies with one site each were not available at the time of analysis, and one of the two-site control groups used only one control site. Seven of the programs were co-located with their corresponding experimental ACT programs.

Data Collection. All programs were rated, using the DACTS, between August 1995 and January 1996. The measure was distributed to principal investigators in the ACT and VA IPCC research studies, along with guidelines specifying data sources and potential issues in definition and interpretation. Formal study periods for the VA and some of the ACT programs had recently concluded, so ratings of these and corresponding control programs were made retrospectively. However, each research team was very familiar with its programs on the basis of extensive data collection and observation of both experimental and control groups, and principal investigators had reacted to earlier drafts of the DACTS. Final ratings of the VA IPCC programs were made by the central research team on the basis of extensive data collection and observation, as well as provisional ratings and other information from site managers. Ratings for ACCESS sites were made by lead evaluators within each state, using preliminary ratings based on structured interviews with selected informants from each site. Informants varied in their familiarity with the content of specific items so, in making final composite ratings, raters weighted their responses accordingly. ACCESS service programs had been fully implemented for at least one year; thus, all the programs in the current study had matured at least beyond the initial start-up phase. Data for all programs except ACCESS were sent to the first author on paper forms.

Analysis. Descriptive statistics were generated for each of the variables to evaluate the distribution of scores. Data from the VA IPCC programs did not include scores on substance abuse treatment variables, so analyses including these variables were performed only on the other three groups. Because multiple ratings were available for only one set of programs, no attempt was made to assess agreement across data sources. Principal components analysis with varimax rotation was used to identify the structure of variation in this dataset; although the ratio of cases to variables was approximately 2:1, the method is appropriate for this kind of descriptive purpose. Variables loading primarily on each factor were averaged to create scale scores. Correlations among scales were calculated, and Cronbach’s alpha coefficients were calculated for scales, for the three a priori dimensions, and for the overall scale to assess internal consistency. Program groups were compared on all scales using analysis of variance and Tukey HSD tests with alpha set to .05, a relatively liberal criterion appropri-
ate for exploratory purposes (Huberty & Morris, 1989). Hierarchical, agglomerative cluster analysis, employing the centroid method for calculating linkage distance, was performed on all programs, using factor scores as variables. Resulting clusters were evaluated by the authors and contributing researchers on the criterion of face validity, based on familiarity with program characteristics. All analyses were conducted using SPSS (Norusis, 1994).

RESULTS

Item Scores

Item distributions were examined to determine their utility in this population of programs. Table 3 shows means, standard deviations, skewness, and number of rated programs for the 26 items used in this study. The majority of items showed appropriate distributions for the sample. All but two items used the full range of the 5-point scale; the two exceptions were caseload size and staffing capacity, for which no programs were rated at the lowest value of 1. All means but three were between 2 and 4; the exceptions were that most of the programs maintained close to full staffing capacity, had a low intake during the period rated, and did not have their own vocational staffing. All standard deviations but two were greater than 1; the exceptions were lower variance on staffing capacity and on the dual diagnosis model. Aggregated across all items, percentages of values assigned on the 5-point scale were as follows: 1=13%, 2=17%, 3=23%, 4=29%, 5=18%.

To measure the degree to which the overall measure tapped an empirically coherent construct of ACT, Cronbach’s alpha was calculated for all 26 items. As shown in the overall scales section of Table 4, the value for the 33 programs reporting all items was .92. Seven items had item-total correlations of approximately .4 or less. When these were removed, the resulting value of Cronbach’s alpha for the 19 remaining items increased to .94. The 19 items are identified on Table 3. However, it is not evident that ACT is necessarily an intrinsically single construct. Because the intent of the current study was to explore variation, if any, beyond a single construct, and because each item was designed to tap an important and substantive aspect of the service model, all variables were included in subsequent analyses.

Factors and Scales

Three a priori scales were calculated as the mean of included items. Table 4 shows
Cronbach’s alpha for each, as well as correlations among them. As expected, the scales are not strongly coherent or differentiated. Internal consistency is only moderately high, given the number of included variables; it is constrained for each scale by one or more variables with low item-to-total correlations. Additionally, interscale correlations are relatively high, suggesting that the variation in this sample is not structured around these three a priori dimensions.

Accordingly, principal components analysis was used for exploratory data reduction. Using an eigenvalue limit of 1.0, eight factors were identified, explaining 78% of the variance. Variance explained by the first unrotated factor, a measure of the strength of the central construct, was 36%. Varimax rotation yielded factors with satisfactory construct coherence. Because each of the last three factors had fewer than three variables with high loadings, other solutions with smaller numbers of factors were derived. Alternative solutions were examined for general conceptual consistency within factor and discrimination across factors. However, the initial 8-factor solution demonstrated the best consistency and discrimination overall. Since factor analysis is used

Table 4
SCALE INTERNAL CONSISTENCIES AND CORRELATIONS BETWEEN SCALES

<table>
<thead>
<tr>
<th>SCALE</th>
<th>ITEMS</th>
<th>N</th>
<th>CHRON α</th>
<th>SCALE CORRELATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Scales</td>
<td>26-item</td>
<td>26</td>
<td>33</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>19-item</td>
<td>19</td>
<td>33</td>
<td>0.94</td>
</tr>
<tr>
<td>A Priori Scales</td>
<td>1. Human Resources</td>
<td>10</td>
<td>49</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>2. Organization Boundaries</td>
<td>7</td>
<td>48</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>3. Services</td>
<td>9</td>
<td>34</td>
<td>0.83</td>
</tr>
<tr>
<td>Empirically Derived Scales</td>
<td>1. Team/Intensity</td>
<td>5</td>
<td>49</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>2. Community Treatment</td>
<td>6</td>
<td>48</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>3. Engagement/Retention</td>
<td>4</td>
<td>49</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>4. Substance Abuse Tx</td>
<td>3</td>
<td>34</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>5. Specialized Staffing</td>
<td>3</td>
<td>50</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>6. Caseload Distribution</td>
<td>2</td>
<td>50</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>7. Staffing/Continuity</td>
<td>2</td>
<td>50</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>8. Vocational Specialist</td>
<td>1</td>
<td>50</td>
<td>—</td>
</tr>
</tbody>
</table>

Table 5
FACTORS, PERCENT OF TOTAL VARIANCE, AND PRINCIPAL ITEM LOADINGS

<table>
<thead>
<tr>
<th>FACTOR</th>
<th>VAR.</th>
<th>LOADING</th>
<th>FACTOR</th>
<th>VAR.</th>
<th>LOADING</th>
</tr>
</thead>
<tbody>
<tr>
<td>Team &amp; Intensity</td>
<td>14.2%</td>
<td></td>
<td>Substance Abuse Treatment</td>
<td>10.5%</td>
<td></td>
</tr>
<tr>
<td>Contact frequency</td>
<td>0.82</td>
<td></td>
<td>Substance abuse groups</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>Team meeting</td>
<td>0.68</td>
<td></td>
<td>Substance abuse treatment</td>
<td>0.78</td>
<td></td>
</tr>
<tr>
<td>Team approach</td>
<td>0.64</td>
<td></td>
<td>Dual disorder model</td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td>Service intensity</td>
<td>0.64</td>
<td></td>
<td>Specialist Staffing</td>
<td>9.0%</td>
<td></td>
</tr>
<tr>
<td>Small caseload</td>
<td>0.63</td>
<td></td>
<td>Substance abuse specialist</td>
<td>0.78</td>
<td></td>
</tr>
<tr>
<td>Community Treatment</td>
<td>13.5%</td>
<td></td>
<td>Nurse on staff</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>In-vivo services</td>
<td>0.68</td>
<td></td>
<td>Psychiatrist on staff</td>
<td>0.58</td>
<td></td>
</tr>
<tr>
<td>Hospital admission responsibility</td>
<td>0.63</td>
<td></td>
<td>Caseload Distribution</td>
<td>7.4%</td>
<td></td>
</tr>
<tr>
<td>Treatment responsibility</td>
<td>0.62</td>
<td></td>
<td>Team leader role</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>Admission criteria</td>
<td>0.61</td>
<td></td>
<td>Intake rate</td>
<td>0.63</td>
<td></td>
</tr>
<tr>
<td>Crisis services</td>
<td>0.61</td>
<td></td>
<td>Staff Capacity &amp; Continuity</td>
<td>7.1%</td>
<td></td>
</tr>
<tr>
<td>Work with supports</td>
<td>0.58</td>
<td></td>
<td>Staff capacity</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>Engagement/Retention</td>
<td>10.8%</td>
<td></td>
<td>Continuity of staffing</td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td>No-drop policy</td>
<td>0.79</td>
<td></td>
<td>Vocational Specialist</td>
<td>5.7%</td>
<td></td>
</tr>
<tr>
<td>Assertive engagement</td>
<td>0.73</td>
<td></td>
<td>Vocational specialist</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td>Hospital discharge planning</td>
<td>0.69</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time-unlimited service</td>
<td>0.57</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
here for its heuristic and descriptive value, these are the results presented.

Table 5 shows the eight factors and variables, with primary loadings for each factor. Across all scales, these loadings ranged from .57 to .90, with a mean of .70. The first, second, third, and fifth factors, together accounting for almost half of the variance, covered most of the salient features of ACT. The team/intensity factor included items measuring shared caseload and team meeting frequency, as well as other items shown. The community treatment factor included a structure and range of service types characteristic of the ACT model; in addition to other items, this factor covered rehabilitative and support services provided in the community and a clearly defined service population. The engagement/retention factor included a demonstrated commitment to ongoing services, as well as actions to develop and maintain engagement over time. The specialist staffing factor included the presence of the listed disciplines or forms of expertise on the team.

The substance abuse treatment factor included three items defining specific aspects: use of specialized groups, amount of time-unlimited treatment, which loaded primarily on engagement/retention, also loaded negatively on community treatment and positively on caseload distribution. Psychiatrist on team, which loaded primarily on specialist staffing, also loaded on team/intensity. Finally, work with support system, which loaded

<table>
<thead>
<tr>
<th>SCALE</th>
<th>PROGRAMS</th>
<th>COMPARISONS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ALL (N=50)</td>
<td>ACT (N=14)</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26-item</td>
<td>3.38 (0.69)</td>
<td>4.01 (0.39)</td>
</tr>
<tr>
<td>19-item</td>
<td>3.46 (0.79)</td>
<td>4.17 (0.45)</td>
</tr>
<tr>
<td>A Priori</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human Resources</td>
<td>3.40 (0.73)</td>
<td>4.00 (0.44)</td>
</tr>
<tr>
<td>Organiz. Bounds</td>
<td>3.59 (0.80)</td>
<td>4.32 (0.46)</td>
</tr>
<tr>
<td>Services</td>
<td>3.20 (0.72)</td>
<td>3.79 (0.40)</td>
</tr>
<tr>
<td>Empirically Derived</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Team/Intensity</td>
<td>3.32 (0.95)</td>
<td>4.33 (0.45)</td>
</tr>
<tr>
<td>Community Tx</td>
<td>3.33 (0.89)</td>
<td>3.81 (0.60)</td>
</tr>
<tr>
<td>Engage/Retention</td>
<td>3.63 (0.91)</td>
<td>4.44 (0.61)</td>
</tr>
<tr>
<td>Subst. Abuse TxC</td>
<td>2.95 (0.99)</td>
<td>3.36 (0.66)</td>
</tr>
<tr>
<td>Specialized Staff.</td>
<td>3.14 (1.29)</td>
<td>4.02 (0.96)</td>
</tr>
<tr>
<td>Caseload Distrib.</td>
<td>4.07 (1.03)</td>
<td>4.71 (0.47)</td>
</tr>
<tr>
<td>Staff/Continuity</td>
<td>4.12 (0.76)</td>
<td>4.32 (0.99)</td>
</tr>
<tr>
<td>Vocational Spec.</td>
<td>1.78 (1.22)</td>
<td>1.86 (1.35)</td>
</tr>
</tbody>
</table>

$a$Mean scores are given, with standard deviations in parentheses.
$b$Tukey's HSD test (.05)
$c$Overall N for Substance Abuse Treatment Scale=37.
primarily on community treatment, also loaded on vocational specialist.

Table 4 shows Cronbach’s alpha for each empirically derived scale as well as correlations among them. The first five scales showed satisfactory internal consistency, with scores of .74 or higher. The next two, both 2-item scales, had lower but usable scores of .64 and .65. Correlations between all pairs of scales except one were .56 or less (i.e., shared variance of 31% or less), typically much less, with a mean correlation among all scales of .32. Team/intensity and community treatment were most strongly correlated with other scales and especially with each other.

Comparison Across Program Types

The four different program types were compared on all scales. Table 6 shows scale means for the whole sample and for each program type. Overall fidelity-to-ACT scores, based both on the total 26 items and on the 19 items with item-total correlations above .4, are shown in the first section. Respective values using the two scales varied slightly, with the leaner 19-item scale yielding more extreme scores, but the results were equivalent. ACT research sites showed higher overall fidelity scores than all other programs, and control programs showed the lowest. Homeless and VA program types were not significantly different from each other on these scales.

On the a priori scales shown in Table 6, the difference between ACT and VA programs on the services scale was just marginally nonsignificant with the statistical test used. Otherwise, consistent with their high intercorrelations, these scale scores followed the same pattern across program types as the overall fidelity scales.

Six of the eight empirically derived scales (see Table 6) showed significant between-type variance. The first, team/intensity, sharply discriminated among all four program types. The next five scales discriminated controls from ACT but varied in the way they discriminated among other program types. The VA programs did not differ from ACT on these scales except for substance abuse treatment, on which VA programs were not rated. The homeless programs were similar to ACT except on engagement/retention, on which they were equivalent to controls. In general, all program types had low involvement of vocational specialists and high staffing/continuity, and did not differ significantly on these two scales.

Empirical Typing

Cluster analysis was performed on the 50 programs, using the eight factor scores that had accounted for approximately four-fifths of the variance among the original set of 26 variables. Two cluster solutions had generally good face validity, one with ten clusters, the other with four. Cluster assignments in each solution were shared and discussed among researchers familiar with the programs. There was consensus that similar programs were clustered appropriately at each level. In many instances, there was substantial consistency between cluster assignments and nominal program type. However, programs that were outliers in their types seemed to be appropriately classified with other program types. Since description of the ten-cluster solution is beyond the scope of the present article, consideration is limited to the simpler four-cluster solution.

Table 7 shows the cross-tabulation of type and cluster memberships for the four-clusters. The program n's are shown in the first section. The first, team/intensity, sharply discriminated among all four program types. The next five scales discriminated controls from ACT but varied in the way they discriminated among other program types. The VA programs did not differ from ACT on these scales except for substance abuse treatment, on which VA programs were not rated. The homeless programs were similar to ACT except on engagement/retention, on which they were equivalent to controls. In general, all program types had low involvement of vocational specialists and high staffing/continuity, and did not differ significantly on these two scales.

Table 7 shows the cross-tabulation of type and cluster memberships for the four-
cluster solution as well as overall fidelity scores on the 26-item DACTS for each cluster. On the basis of their respective program type memberships and factor scores, the four clusters were characterized as strong ACT programs, rehabilitation teams, weak ACT programs, and standard case management (CM) programs. The strong ACT cluster was dominated by ACT research programs, but about half of the VA and homeless programs were in this group. These programs had especially high scores on team/intensity, community treatment, specialist staffing, engagement/retention, and staffing/continuity. The rehabilitation team cluster was high on community treatment, substance abuse treatment, specialist staffing, and vocational specialist; this group included programs that were somewhat less intensive and team-oriented, but more strongly focused on community-based vocational interventions. The third cluster was designated weak ACT because it included several programs that had ACT-like structures but had generally achieved lower functioning as teams, having been less successful in hiring and retaining appropriate staff, and client turnover had been high; accordingly, they scored lower on staffing/continuity, specialist staffing, and engagement/retention. The fourth group, standard case management, included all control sites except one that had done well at emulating the ACT intervention, as well as one VA site that had not been successful in implementing the intensive case management program.

**DISCUSSION**

The foregoing sections have presented the development and use of a 26-item anchored measure of fidelity to ACT, the Dartmouth ACT Scale (DACTS). Items represent criteria for model fidelity that were derived theoretically or empirically. Specific indicators and data sources for these criteria were selected in terms of both substantive relevance and feasibility of measurement. Performance of the measure was evaluated by applying it to 50 programs, including programs expected to have high fidelity to ACT, programs with features similar to those of ACT, and standard treatment programs. Issues to be discussed include measurement properties of the scale, variation among programs, potential uses, and possible future development of the measure.

**Measurement Properties**

Findings on the 26 items used to form the DACTS were generally as expected with this set of programs: most variables, as well as the overall scale, showed appropriate means and distributions. Because all but the control programs represented ACT-like models, means somewhat above 3.0, the midpoint of the range, are consistent with the design intent to capture the variance between standard case management and a fully implemented ACT program. Mean scale values for high and low groups reflected overall use of the scoring range without risking floor and ceiling effects.

Further evaluation is necessary to determine modifications appropriate to specific items, e.g., to those defining the last two factors. The vocational specialist variable had a very low mean, was relatively skewed, and was not well correlated with the central ACT construct in this set of programs. However, this single variable played a strong role in discriminating a cluster of ACT-like programs that were especially vocationally oriented. Further, as the norms for effective interventions for people with severe mental disorders evolve, vocational interventions will probably be more salient (Bond, Drake, Mueser, & Becker, 1997). Although there was less variance on staffing capacity and continuity items, this scale played a role in identifying poorly implemented programs. Maintenance of staffing is not a requirement unique to ACT, but it is fundamental to continuity of care, a core feature of the model. These two factors, along with other factors defined by small numbers of items, may require supplementation through addi-
tional variables consistent with the construct.

Support for validity of the DACTS comes from four types of sources. First, the items have generally good face and content validity, explicitly reflecting the essential features reported in an extensive literature about the model. Second, although the Madison PACT program was not included in the sample as a criterion case, trainers from that program have given a generally positive critique of the items and standards set in the measure (D. Allness, personal communication; May, 1996). A similar conclusion was presented in an authoritative review of approaches to operationalizing the model (Meisler, 1997). Third, by using the DACTS for groups of programs with generally known properties, the current study empirically establishes the measure's capacity to function as designed, i.e., to distinguish programs at different levels of fidelity to ACT. Finally, there is some preliminary evidence of predictive validity. In an unpublished study evaluation of 12 newly formed ACT teams, Bond, Salyers, and Fekete (1996) found that scores on the DACTS were associated with reduced hospital use and lower levels of staff burnout. In preliminary analyses by the first and third authors and colleagues, components of fidelity as measured on the DACTS were related to substance abuse outcomes in dual-diagnosis ACT programs. More direct and rigorous evaluation of predictive validity awaits further research.

The empirical support for validity necessarily provides only indirect support for reliability, which was not directly evaluated in this study. Work remains to be done in assessing different types of reliability for the DACTS, including agreement both between data sources and between interviewers collecting the information. A study using a very similar measurement approach found that agreement between interviewers was very high (Bond, Becker, Drake, & Vogler, 1997). On the other hand, agreement between data sources in recently formed teams was found to be lower, at least within a narrow range of program variability (Bond et al., 1996; Test, Bond, & McGrew, 1997).

Variation Among Programs

Factor analysis was used to reveal patterns of relationships among the criteria, i.e., to observe the structure of variation in the implementation of ACT in the sample of programs. The eight factors showed generally good face validity in that primary variables for most of them allowed for clear interpretation. The first three factors, team/intensity, community treatment, and engagement/retention, along with the fifth, specialist staffing, captured most of the core traditional criteria of ACT. The fourth factor, substance abuse treatment, although historically less explicitly associated with ACT, has come to be recognized as an integral treatment component (Stein & Santos, 1998):

As was evident both in factor loadings and in correlations between resulting scales, there is a close association between team/intensity and community treatment. The former covers both the capacity to provide high-intensity services as needed and a team structure that enhances the staff information base, allowing optimal deployment of human resources. These characteristics would appear to enhance significantly a program's prospects for addressing the criteria in community treatment, which covers the range of responsibility and activity necessary to address clients' needs. Engagement/retention includes criteria reflecting the separable issues of how actively clients are engaged and how long they are retained in the program. Assertive engagement and time-unlimited treatment have historically been core concepts for ACT, but in the ACT-like programs in this sample there was greater variation. The ACCESS programs, for example, were mandated to serve a new population of 100 homeless clients each year. Clearly, some form of transition or graduation is necessary in this context.

Two of the remaining three factors were
represented by only two criteria, and one of them by a single criterion. In the normal course of instrument construction, such dimensions would be modified by elimination, elaboration, or integration with other dimensions. They are reported here, however, because the present purpose is partially exploratory and descriptive. Staffing/continuity was internally consistent, interpretable, and useful in differentiating programs. The caseload distribution factor posed a greater challenge in interpretation, and this label, especially, is necessarily tentative. Both of the included criteria were derived from the work of McGrew et al. (1994), in which the role of the team leader as provider of direct service alongside other team members was shown to be a predictor of improved outcome. Although the authors did not explicitly address how this might occur, a possible mechanism is improved organizational functioning and climate, which itself could follow from more egalitarian roles fostered by a leader who is involved with treatment and is not just a distant administrator. A low, controlled intake rate has been proposed as a critical ingredient; this, too, would ensure that the distribution of relationships between staff and clients would be even and comfortable over time. Unlike most of the first five factors, which appear to reflect more traditionally central features of ACT, these last three seem improbable candidates for replication in other studies.

**Utility of the DACTS**

An important use of a fidelity measure is to discriminate programs faithful to a model from those that are not. Both overall and individual scale scores on the DACTS played a role in contrasting the different types of programs included in this sample. Control programs scored below all other programs, both overall and on six of the eight scales. The VA and homeless programs were not designed to replicate the full array of ACT characteristics, and they scored appropriately at intermediate levels of overall fidelity, significantly different from both ACT and controls but not from each other. The team/intensity scale discriminated among all four programs, distinguishing the more team-based and intensive homeless programs from the VA programs. The homeless programs were not designed for long-term retention of clients and consequently scored on a level with controls on the engagement/retention scale, despite commitment to assertive outreach. It should be noted that the criterion for assertive engagement was operationalized in such a way that it emphasized use of legal mechanisms relative to less easily measurable engagement strategies. Many of the homeless programs attempted to minimize use of legal mechanisms, so their lower score on this scale may have been partly an artifact of operationalization.

Although overall and scale scores on the DACTS were useful in discriminating among groups of substantially different programs, the measure was not designed to make fine distinctions within sets of very similar programs. Anchors for the five-point scales were set to the range between conventional services and ACT; sensitivity to differences is lowered when a truncated range must be used. The factor structure reported here was derived by using a sample covering most of the range of measurement; although factors approximating the first five may well emerge in similar samples, the overall structure is not necessarily replicable or relevant in more homogeneous samples. On the other hand, the generally satisfactory classification of programs using cluster analysis, as reported here, suggests that thorough use of the total variance may have some utility beyond the simpler use of scale scores. As noted, the importance of these sources of variation for outcome remains to be examined. Use of the DACTS in studies that include outcome data will be necessary in order to determine its utility in verifying that putatively critical ingredients are indeed critical.

Aside from characterizing differences
among types of programs and evaluating relative contributions to program effectiveness, a fidelity measure such as the DACTS has other important applications. Program implementers and researchers can use it as a reference for ensuring program quality either overall or in terms of each individual criterion. The overall fidelity score can be used as a general indicator of fidelity, perhaps with a specific threshold for inclusion of sites in a particular study. Greater use of this or similar measures would allow comparison with norms for different types of programs; the data reported here may serve as preliminary norms. A number of consultants have reported that the format of the DACTS makes it useful as a training instrument by providing concrete quantitative or qualitative targets for self-evaluation and emulation. This type of measure is also amenable to multiple applications in a given setting. The full set of criteria can be used for implementation monitoring and quality management, while reduced sets of variables such as the more reliable 19-item overall fidelity scale or selected subscales can be used for statistical analysis.

Further Development

The version of the DACTS reported here is a step in the process of operationally defining the critical components of ACT. This evolution is driven by several factors, including more careful explication of existing notions and experience with the model, advances in effectiveness research, and elaboration of program theory. The inclusion of substance abuse criteria represents an instance of elaboration, since this clinical area had not initially been featured as a part of the model. After collecting the data in this study, two items were added: program size, and the role of consumers on the treatment team. The first follows from the experience of the leading ACT training group from Madison, where absolute team size was found to be a critical element in ensuring the necessary coverage (D. Allness, personal communication, May, 1996). Along with the criterion of time-unlimited treatment, the importance of this element is currently subject to debate among proponents of ACT models, but it seems nonetheless an important dimension to track. The second new item, on consumer roles, follows from a gradual increase in attention in the field to the potential benefits of consumers' participation in the direct provision of services. Again, this is an important feature to evaluate across programs, even if it is not currently considered a formal criterion.

Additional modifications to the DACTS have been recommended, including more explicit elaboration of the clinical services themselves. The authors expect to work closely with the principals of PACT, Inc., a project established by the National Alliance for the Mentally Ill to promulgate the ACT model throughout the country (Flynn, et al., 1997). PACT, Inc. is expected to provide both training of staff and certification of programs. This group has developed a detailed training manual (Allness & Knoedler, 1998). In this context, the DACTS may serve as a prototype for the kind of multipurpose measure that such an endeavor would require, supporting training, empirically based fidelity evaluation, and multidimensional process measurement to complement comprehensive outcome measurement. This complete range of applications is necessary for measurement of model fidelity in order to evaluate the quality and quantity of critical dimensions in services.

CONCLUSION

ACT is a complex community-based service approach that has been demonstrated as helpful for people with severe mental disorders adapt to lives in the community. Widespread replication of the model has been attempted. However, without very explicit operational criteria for program structure and processes, developers of new programs cannot be sure that they are following the model accurately. Further, without
careful measurement of the multiple dimensions of this complex program model, researchers can neither be sure that specific programs are true to the model nor identify the relative contribution of program components to effectiveness. Pending better understanding of how specific components of treatment programs and settings interact with specific characteristics of the people involved to produce desired outcomes, researchers must sample from a range of places along the hypothesized causal chain, including both structural elements and intermediate process elements.

The Dartmouth ACT Scale reported here provides a preliminary, standardized measure of fidelity to the ACT model. Empirically and theoretically derived criteria address multiple dimensions of program structure and process. These criteria and the overall measure discriminated among a variety of ACT and ACT-like programs and discriminated these from conventional treatment programs. Although the current measure will be refined to enhance its utility in developing and evaluating ACT programs, it has been useful in its current form for training and fidelity determination.

REFERENCES


Family-aided assertive community treatment: A comprehensive rehabilitation and intensive case management approach for persons with schizophrenic disorders. *New Directions for Mental Health Services*, 53, 43–54.


