A Fidelity Rating Instrument for Consumer-Run Drop-in Centers (FRI-CRDI)

Carol T. Mowbray
Mark C. Holter
Lori Stark
Carla Pfeffer
University of Michigan

Deborah Bybee
Michigan State University

Objective: Given the present emphasis on accountability and maintaining quality, the objective of this study was to develop, apply, and assess the reliability of a fidelity rating instrument for consumer-operated services—a promising model, but one for which fidelity criteria are not yet established. Method: Based on observations, documents, and director interviews from 31 consumer-run drop-in centers, we developed a scale measuring fidelity to pre-established criteria and rated each center on scale items. A second study examined the interrater reliability of the measure. Results: Scale scores on the 31 centers showed substantial heterogeneity on the majority of the criteria. The fidelity rating scale demonstrated satisfactory interrater reliability on most items. Conclusions: The fidelity rating instrument is ready to be used by social work researchers evaluating consumer-operated services. Furthermore, social work researchers can use similar methods with other innovative services that should be evaluated but currently lack standards and fidelity criteria.

Keywords: *fidelity; consumer-operated services; mental illness*

Self-help as an adjunct or an alternative for individuals with addictions (gambling, alcohol, etc.), health conditions (colostomy, cancer, overweight, etc.), or problematic life circumstances (divorce, single parenthood, widowed, etc.) is now accepted as a valuable resource to health and social service systems (Borkman, 1990). For those with serious mental illnesses, self-help is of much more recent origin.

Similar to self-help, in terms of its values, are consumer-run services (CRS). CRS encompass a variety of program types in which the services are provided by psychiatric consumers to their peers. They differ from self-help in that the CRS is a formal program with a governance structure and operates as a business to provide these services. In CRS, the paid staff and at least a majority of the governing board are themselves consumers. The programs are funded by state or local mental health authorities and are contracted to provide a specific set of services.

Many professionals and researchers have recognized the potential benefits of CRS, especially in light of the expanding consumerism and recovery movement among those with psychiatric disabilities, viz., ". . . this model of community-based services . . .is an important new direction in community care and one that may provide important opportunities for social work collaboration . . ." (Test, 1998, p. 429). Consumer and nonconsumer authors have articulated the benefits of CRS, such as sharing similar life experiences and offering a different worldview to assist in making sense of experiences (Davidson, Chinman & Kloos, 1999); personal control (Salem, 1990) that counteracts typical feelings of powerlessness (Chamberlin & Rogers, 1990); offering choice (Connelly, Keels, Kleinback, Schnieder, & Cobb, 1993); promoting independence and competence, providing social support, and individualized services (Chamberlin, 1984); providing a unique and needed support that is

Authors' Note: This research was supported through a National Institute of Mental Health grant to the University of Michigan, School of Social Work, Center for Research on Poverty, Risk, and Mental Health. Correspondence may be addressed to Carol T. Mowbray, Ph.D., Center for Research on Poverty, Risk, and Mental Health, School of Social Work, University of Michigan, 1080 South University Avenue, Ann Arbor, MI 48109-1106 or via the internet using cmowbray@umich.edu.

Research on Social Work Practice, Vol. 15 No. 4, July 2005 278-290
DOI: 10.1177/1049731505275060

more empathic, concrete, and relevant (Mowbray, 1997); offering role models and organizational involvement (through a flattened administrative hierarchy; Segal, Silverman, & Temkin, 1993); instilling hope (Kennedy, Humphreys, & Borkman, 1994) to facilitate recovery (Yanos, Primavera, & Knight, 2001); and working toward social justice and social change on behalf of individuals with serious mental illnesses (Chamberlin & Rogers, 1990; Segal et al., 1993). Policy and programmatic initiatives at state and federal levels have supported consumer involvement in service planning and provision; some have even mandated incorporation of CRS in mental health systems through programmatic recommendations and requirements (Von Tosh & Delvecchio, 1998).

A major problem with research on and expanded development and promotion of CRS is a lack of clarity concerning what the essential ingredients are, along with interchangeable use of terms like *consumerism, consumer-involved, consumer-run*, and *self-help*. Segal et al. (1993) concluded that even the basic practices of consumer-run agencies are poorly documented and poorly understood. Mowbray, Moxley, Jasper, and Howell (1997) presented a framework for organizing consumer roles as providers of mental health services. The major dimensions of the framework are: Who has control of the services (consumers vs. nonconsumers), and what is the aim of the alternative service (mutual support vs. formal service provision)? Within this two by two matrix, CRS reflect an aim of formal service provision controlled by the consumers themselves; in contrast, self-help groups (like the Bipolar and Depression Support Alliance, Schizophrenics Anonymous, or GROW) are consumer-controlled but aim to provide mutual support—not formal services. CRS and self-help organizations also have differing organizational structures—the former usually is a free-standing legal entity with its own Board, formal resources, and specifically defined service provision.

As might be expected, given the lack of definitional clarity for CRS, there is considerable heterogeneity among programs identified as consumer-run. For example, in a statewide survey of Michigan's consumer-run programs (specifically, the drop-in center model), Mowbray, Robinson, and Holter (2002) found that the number of paid staff positions per center ranged from zero to eight, with a median of three; and that the annual budgets varied widely, from a low of $2,080 to a high of $257,000 (median $54,000). Variation in services provided has also been documented. Chamberlin, Rogers, and Ellison (1995) reported common activities at six consumer-run drop-in (CRDI) centers studied, which included assistance with legal problems and with housing; but in the Michigan study (Mowbray et al., 2002), legal assistance was not mentioned and housing assistance was offered in only 12% of the centers. However, similar to the Chamberlin et al. (1995) survey, Michigan's CRDI centers did focus on social or recreational activities, transportation, help in finding jobs, and provision of basic needs (for food and clothing; Mowbray et al., 2002).

The heterogeneity of consumer-run programs undoubtedly reflects their grassroots origins and the barriers they often face in start-up, such as lack of funding and support from traditional mental health providers. Other prominent service models—including assertive community treatment, supported employment, and clubhouse—have had the advantage of being based on a research demonstration project or a specific set of standards from which the elements and operation of the model could be described and measured, serving as a check for further replications and for assessing quality across different contexts and with different target populations. CRS do not have any specifically identified prototype that has been operated. The service model has been primarily articulated in the writings of consumer leaders and advocates. There are no program manuals to provide guidance to those wanting to replicate a CRS model nor to those desiring to carry out research on CRS who need to be assured that the program they are assessing conforms to a CRS model.

Such heterogeneity of operations is problematic for multiple reasons. Evaluation research has established that interventions that adhere to prescribed standards usually produce better outcomes (McHugo, Drake, Teague, & Xie, 1999). Furthermore, without specified standards and criteria, replications of program models are difficult; they may drift and lose their fidelity to intended values and principles, often reverting to traditional program operations (Bond, Williams, & Evans, 2000). The inability to measure model adherence is also a serious disadvantage in conducting outcome studies, as well as in interpreting their results, as lack of success may reflect failure of the model or failure to implement the model as intended (Orwin, 2000).

To address the often observed heterogeneity of programs that are supposed to be following a given model, fidelity criteria have been established for some program types. Fidelity has been defined as "the adherence of actual treatment delivery to the protocol originally developed" (Orwin, 2000, p. S310). According to Bond, Evans, Salyers, Williams, and Kim, (2000), "fidelity refers to the degree to which a particular program follows a program model . . .a well-defined set of interventions and procedures to help individuals achieve some desired

goal" (p. 75). For administrative purposes, fidelity criteria can be used as a guide to implementing a program following a specific model (Bond, Becker, Drake, & Vogler, 1997), or to monitoring programs to help assure quality (see Bond, Williams, et al., 2000, for examples). Having fidelity criteria can also promote external validity by providing adequate documentation and guidelines for replication projects adopting a given model. In treatment effectiveness research, fidelity criteria are used as a manipulation check to ensure internal validity (Hohmann & Shear, 2002). Several authors (Bond, Evans, et al., 2000; Brekke & Test, 1992) have noted that fidelity criteria may be especially needed in the mental health field as programs often lack model specification and model adherence and rely extensively on clinical knowledge and skill.

Recent research has shown the usefulness of quantifiable measures of fidelity criteria. Fidelity measures have been able to differentiate programs following a specific treatment model versus standard approaches for two areas: assertive community treatment (vs. standard case management; Teague, Bond, & Drake, 1998) and the individual placement and support model of supported employment (vs. traditional vocational rehabilitation programs; Bond et al., 1997). Furthermore, clients in high-fidelity versus low-fidelity versions of model programs have been found to achieve significantly more positive outcomes (Blakely et al., 1987; McHugo et al., 1999; McGrew, Bond, Dietzen, & Salyers, 1994) and programs that pass versus fail a measure of program fidelity were judged to have superior performance (Macias, Propst, Rodican, & Boyd, 2001).

The purpose of this article is to describe our process of developing a fidelity rating scale, applying it, and measuring interrater reliability for CRS, specifically CRDI centers, which are perhaps the most common model of CRS. These CRDI centers are intended to serve critical social support and socializing functions, providing organized and informal recreational and social activities where participants and staff (all consumers) can assist each other in solving their daily living problems. CRDI centers should provide a safe, supportive, and normalizing environment in the community for individuals labeled *mentally ill*—especially those who are isolated and not regular participants in traditional mental health programs. These centers offer an atmosphere of acceptance where individuals can feel needed and grow in self-worth, dignity, and self-respect, and where they can learn from each other about resources and services available and how to access them.

In this article, we first provide some background on how the fidelity criteria were identified. We then describe our process of developing a fidelity rating scale for CRS and applying it in a statewide sample of CRDI centers. We also present results from a second study, measuring interrater reliability of the scale in a separate sample of CRDI centers. Finally, we discuss the potential use of this instrument for research and quality control.

BACKGROUND

In recognition of the well-accepted admonition, "nothing about us without us," this study was conducted in collaboration with the Justice in Mental Health Organization (JIMHO), a statewide consumer organization that provides technical assistance, training, and support to consumer-run service programs. JIMHO was a partner in developing the study design and applying for federal funding from the National Institute of Mental Health. JIMHO recognized that future funding for consumer-operated services might well be contingent on establishing the effectiveness of the CRS model. They supported the need for fidelity criteria and, ultimately, for outcome studies. JIMHO's involvement was supported with grant funds, through a contract with the University.

The development of the fidelity scale on CRDI centers followed approaches recommended in the literature. We specified the dimensions of this consumer-run service model as well as those reflecting the structure and activities of the program and the behavior of the staff (as opposed to opinions, attitudes, or outcomes; Bond, Williams, et al., 2000). We attempted to make scale items objective (McGrew et al., 1994), and we included key process variables, such as client choice (Paulson, Post, & Herinckx, 2002). We also used a multimodal approach to collecting data to assess fidelity (Bond, Evans, et al., 2000). The steps underlying our approach are further described below.

The first step in developing the fidelity rating scale was identifying the criteria for CRS through an extensive literature review of published and unpublished writings of consumers. Criteria identified were grouped into domains for conceptual clarity, following Donabedian's (1980) classic framework of structure and process variables. *Structure* refers to "the relatively stable characteristics of the providers of care, of the tools and resources they have at their disposal, and of the physical and organizational settings in which they work" (Donabedian, 1980, p. 81). Examples of structure criteria include: being

consumer-operated, voluntary attendance and participation, consumer determination of policies and operations, availability (stable location and predictable hours), accessibility, and providing services to meet consumer needs.

Process is the second major dimension described by Donabedian and refers to methods of delivering services. Because the process domain encompasses numerous values and identified activities of CRS, we used concepts developed by Maton and Salem (1995) on the characteristics of empowering community settings for further categorization. The Maton and Salem (1995) approach was chosen because writings of consumers frequently specify empowerment as an intended outcome of participation in CRS. Thus, the process domain was subdivided into:

1. Belief systems—group empowerment, recovery orientation, and personal growth and development.
2. Opportunity role structure—consumers' active participation in operating the center, regular attendance, choices and decision-making opportunities available, opportunities to practice and improve skills (e.g., in communication, interpersonal relationships, or work-relevant areas), and nonhierarchical structure between staff and consumers.
3. Social support—development of social networks, socializing opportunities, feeling a sense of community, self-help, reciprocity, seeing other consumers as positive role models, etc.

Our preliminary fidelity criteria were reviewed by a group of national experts on consumerism ($n = 67$), identified through an extensive search of consumer newsletters (e.g., *The Key*), professional journal articles, book chapters, and conference presentations. Experts were primarily consumers or advocates but also included service providers and researchers; the experts were from 21 states, representing all regions of the country. In Wave I, experts were mailed a survey that asked them to indicate whether each criterion was critical to a CRDI center. Responses were received from 47% of those surveyed ($n = 29$). The experts indicated that all the criteria were critical; one item was reworded for clarification. In Wave II, using the results of Wave I, a Web-based survey was developed in which respondents were forced to rank order the criteria from most to least essential. Respondents rated highest those structural and process components emphasizing the value of consumerism: consumer control, consumer choices and opportunities for decision-making, voluntary participation, and respect for members by staff (Holter, Mowbray, Bellamy, MacFarlane, & Dukarski, 2004).

The present article reports on two studies: (a) our development of an instrument to rate fidelity to the criteria, in which we describe the process of producing the instrument and the resulting ratings, based on 2-day site visits to a statewide sample of CRDI centers; and (b) the results of an interrater reliability study, involving new raters, new training for the raters, and data collection at four additional CRDI centers. Both studies were approved by the University of Michigan Institutional Review Board as well as by Human Subjects Committees at all Community Mental Health agencies that had such review requirements.

STUDY 1: DEVELOPMENT OF THE FIDELITY RATING INSTRUMENT

Methods

Sample and procedures. The purposive sample was composed of 31 CRDI centers across Michigan—18 (58%) urban areas and 13 (42%) rural. Sample selection criteria required the CRDI be in operation at least 2 years; the longest length of time in operation was 23 years, and the mean was 10 years. Number of staff ranged from 1 to 14, with a mean of 3.84 and a median of 3. Number of members attending on a typical day ranged from 5 to 103, with a mean of 30; 65% of consumers at the CRDI centers visited were male, and ages ranged from 18 to 81, with a mean age of 44 years.

The 3-person field research team consisted of the field research director—clinically trained, with several years' experience conducting interviews with adults with psychiatric disabilities—and two research assistants who had prior work experience and training involving people who have mental and emotional problems. The 31 two-day site visits were conducted between June 2001 and August 2002. Centers had agreed to participate in the data collection and site visit dates were arranged by JIMHO, the collaborating consumer organization. The team arrived at each site at a prearranged time and introduced themselves to the director, who had spoken with a team member by phone prior to the visit. Soon after arrival, the field research director gave a short, informal, friendly introduction to the group of consumers at the CRDI to provide an overview of the purpose of the visit and to put consumers at ease with our presence as well as with the procedures of the data collection. At each CRDI center, all consumers present at any time during the 2-day site visit were asked to complete a questionnaire—either self-

administered or as an interview with one of the research assistants—and were paid $5 for their participation. Centers each received $100 compensation.

The field research director interviewed the center director at each site and asked all available CRDI staff to complete questionnaires. During the remainder of the site visit, the field team spent time making qualitative observations from conversations directly with consumers, observing interactions between staff and consumers and of consumers with each other, and observing the physical environment of the interior and exterior of the center, including the layout and condition of the space, postings, furnishings, equipment, and amenities (e.g., washer and dryer, food, showers, etc.). The Instrument for Site Observations (ISO) was devised for the site visits to guide the observations and other qualitative data collected during each site visit.

Following all site visits and data collection, the field research team developed the fidelity measure—the Fidelity Rating Instrument for Consumer-Run Drop-In Centers (FRI-CRDI). The FRI-CRDI is based on the criteria previously endorsed by consumerism experts. Following the approaches of Teague et al. (1998) and as recommended by Bond, Williams, et al. (2000), we used a benchmark process to operationalize each criteria. To do this, the instrument specified indicators—aspects of the structure or process of operations at the CRDI that were most relevant for a particular criterion. The indicators were intended to function as a guide to the raters as to what staff observations or other sources of data should be reviewed to produce a rating on the particular criterion of the FRI-CRDI. For example, these are the indicators for the criterion "Facilitating Referrals":

Are staff knowledgeable about necessary services available to consumers and do they seek to link consumers with these (e.g., affordable housing, transportation, job placement and training, education, food pantries and soup kitchens, clothing distribution sites, Medicaid and Medicare, social security benefits, self-help and support groups, shelters, low income resources, social recreational opportunities, etc.)? Do we hear any talk about the center linking consumers with services? Does it seem that the center has relationships with other helpful agencies and services? Are there updated and accessible pamphlets, postings (including specific, local contact information)? Do representatives from local service programs come to the center? More important than posted information is evidence that staff are knowledgeable, helpful, accessible, and proactive regarding linking consumers with services they need in the community. Does it seem that consumers can become more knowledgeable about available services by hearing the talk at the center about services that are available in the community?

A rating scale was developed for each criterion, using either three (1, 3, 5) or five points (1, 2, 3, 4, 5) reflecting an operation of the combined indicators. The rating scale for each criterion item was intended to be specific, objective, and mutually exclusive and still broad enough to encompass the significant heterogeneity encountered during the site visits. See Table 1 for a list of the fidelity criteria and some of the scale point anchors. It should be noted that standardized scores and other statistics (e.g., correlations, *t* tests) computed from the data are not affected by the rating scale having three versus five points.

The FRI-CRDI was then used by the field research team, as a group, to rate several of the programs, using recordings from the ISO. Following refinement of the rating process, team members rated several other programs individually and then compared ratings. Because individual raters had differing experiences with the programs, it was decided that the ratings of the 31 programs on each criteria should be by consensus, with refinement of the FRI-CRDI along the way. This was an iterative process, whereby deficiencies in the rating instrument were discovered as more centers were rated; the rating instrument was then refined, and raters went back to rerate programs according to the refined instrument.

Results

Table 2 presents the finalized ratings on each FRI item, across the 31 CRDI centers. The five criteria with the highest scores were (a) voluntariness, (b) consumer choice and decision-making, (c) sense of community, (d) general respect, and (e) social support. The greatest number of average high scores was in the *social support* domain; there were no highly rated criteria, on average, in the *process and belief systems* domain. By contrast, the lowest rated items, on average, were *instrumental activities and services* (housing, transportation, education, and job assistance), *practice and improve social and work-related skills*, *social-recreational activities and services*, *consumer involvement*, and *group empowerment*. None of these low-rated criteria were in the *social support* domain; two were in the *belief systems* domain.

TABLE 1: ***Fidelity Rating Criteria for Consumer-Run Drop-in Centers***

Criteria Item	*Lowest Rating (1)*	*Middle Rating (3)*	*Highest Rating (5)*
Structure			
Voluntary—extent that attendance or participation is required or coerced, and members can come and go at will	More than two thirds of consumers compelled to attend; generally can't leave without permission	20% to 35% compelled to attend and participation may be expected; or > 20% compelled but consumers pressed to participate	Attendance and participation completely voluntary; members come and go as they wish and participate only if they choose
Consumer determination of policy, operations, and planning	No consumer board; staff makes virtually all decisions; consumers have almost no input	Director, staff, or small group of consumers make most decisions; little democratic participation	Democratic process is usual and expected: regular open meetings; ideas encouraged and implemented
Transportation—accessibility, affordability, dependability, safety	Difficult for consumers to get to the center: no public transport, remote location, etc.	Public transport unreliable, costly, dangerous, or limits flexibility	Easy for consumers to come and go as they wish
Exterior physical environment—appearance, neighborhood safety, wheelchair accessibility	Neighborhood appears unsafe; exterior unappealing; may not be wheelchair accessible	Neighborhood and center exterior somewhat unpleasant; may be difficult for wheelchairs	Neighborhood and building appear safe and pleasant; entrance is wheelchair accessible
Interior physical environment—space, smokiness, cleanliness, comfort level, wheelchair accessible interior	Potentially unhealthy: smoky, filthy, foul odors, disrepair, crowded, broken furniture, wheelchair inaccessible	Unpleasant: unkempt, lacks comfortable furniture, minor maintenance problems, kitchen too small, or wheelchair barriers	Clean, comfortable; well laid-out; large enough for growth; large enough kitchen; wheelchair accessible
Facilitating referrals—getting member needs met in the community	No useful postings or referrals by staff for necessary services in the community	Staff not proactive v/v SPELL OUT information; helpful if asked; some postings but may be outdated or not user-friendly	Much usable information posted; staff are knowledgeable, helpful, and proactive in sharing information regularly
Outreach to recruit new members and to increase visibility in community	No emphasis on recruiting new consumers or increasing visibility in the community	Center asks case managers at mental health agency to tell new clients about center	Center has mental health agency tell new clients about center, maintains contacts with other local agencies
Activities and services—provision for meeting basic needs in the center	Center provides 0 to 1 of (a) quality meals or snacks; (b) meals or snacks for all almost every day; (c) telephone; (d) washer and dryer; (e) at least one of: hygiene products, clothing, showers, or food pantry	Center provides any three of (a) quality meals or snacks; (b) meals or snacks for all almost every day; (c) telephone; (d) washer and dryer; (e) at least one of: hygiene products, clothing, showers, or food pantry	Center provides all of (a) quality meals or snacks; (b) meals or snacks for all almost every day; (c) telephone; (d) washer and dryer; (e) at least one of: hygiene products, clothing, showers, or food pantry
Housing, transportation, education, and job assistance services	Center provides none of these services on a regular basis	Center provides one of these services regularly	Center provides two or more of these services regularly
Social recreational activities	Center may have TV, stereo, or even a few cards or games but nothing more for in-house activities, and activities away from the center are less than monthly, if any	Center may have excellent social-recreation options in or outside the center, but not both; or center has infrequent activities outside and a few inside beyond TV	Center has many enjoyable options at the center and activities away from the center more than two to three times per month
Process—Belief Systems			
Group empowerment—opportunities (groups, activities, conversations, posted information, conferences) to learn about social and political issues affecting consumers	Consumers see troubles as individual rather than group-based; no lobbying, attending rallies, consumer conferences, or antistigma activities	Some evidence: talk, postings, meetings, events where consumers identify as part of affected group; center may encourage conferences, lobbying, antistigma activities—consumers less interested	Much evidence that consumers identify as members of an affected group: talk about laws, bureaucracies, discrimination, and how to effect change; attend conferences, rallies, lobbying events

(continued)

TABLE 1 (continued)

Criteria Item	*Lowest Rating (1)*	*Middle Rating (3)*	*Highest Rating (5)*
Process—Belief Systems			
Practice and improve social and work-related skills—opportunities provided	No computers, clerical tasks, activities to plan, discussion groups, or meal preparation opportunities	Groups at center plus two of following: (a) computers, (b) clerical tasks, (c) meal prep, (d) info gathering, (e) organize activities, or (f) attend conferences	Groups at center, and three of following: (a) computers, (b) clerical tasks, (c) meal prep, (d) information gathering, (e) organize activities, or (f) attend conferences
Recovery orientation, personal growth, and development—emphasize strengths, skill-building, and independence	Staff lack hope for consumers; sense that consumers should keep expectations low; staff do for rather than with consumers	Perspectives are mixed: one half recovery and hope, one half disability and resignation; or recovery orientation is absent	Hope is pervasive; activities and talk regarding jobs, housing, and education focus on strengths, skills, and independence
Process—Opportunity Role Structure			
Consumer involvement in wide variety of tasks to operate the center	If consumers are involved, it's mainly in janitorial tasks	Some consumers are involved in more than janitorial tasks	Many consumers are involved at varying responsibility and skill levels
Consumer choice and decision-making regarding how to spend their time at center	No choice; participation tightly structured; strict or arbitrary rules	Some choice; some activities may be restricted; may have too many rules	Freely choose level of participation at all times
Nonhierarchical structure between staff and members	Staff maintain strict hierarchy; restrictive rules for consumers, not staff; staff have more resources; staff are condescending	Some hierarchies apparent: some staff condescend, some staff more respectful and encourage democratic participation	Consumers do not automatically defer to staff; no differences in rules or access to resources; staff are not condescending
Process—Social Support			
Member retention activities—friendly atmosphere, orientation for newcomers	Indifferent atmosphere; little or no recognition of consumers' contributions; little welcoming and orienting of newcomers; no contact with those who've been away	Somewhat friendly atmosphere; some recognition of contributions; occasional contact with those who've been away; newcomers introduced to at least two others	Very friendly, welcoming; recognition for contributions; visit or telephone members in hospital and sometimes those who've been away; comprehensive orientation
General respect toward members—polite, respectful, no threat of commitment	General disrespect, unkind, impolite; staff may threaten to report consumers to case managers or others or threaten commitment or treatment	Occasional put-downs by staff or other consumers in a half-joking way; a few get picked on; no threats to report or commit	Pleasant and respectful; disrespect doesn't go unnoticed: others respond to uphold respect as community standard; no threats to report or commit
Respect for diversity—inclusive, no racist, sexist, antigay, or demeaning speech or behavior	Racist, sexist, antigay, or other demeaning speech or behavior is frequent or severe and incurs no meaningful response	Infrequent and less severe racist, sexist, antigay, or demeaning speech with half-hearted or insufficient response	Attitudes and behavior appear free of prejudice; concerted effort toward respectful and inclusive environment
Social support—social relationships and social networks at the center	Superficial relationships; don't know each other well; may be isolated, indifferent, or unfriendly	Meaningful friendships; know about each other's lives and express support	Meaningful friendships that extend beyond the center; or at the center, show extraordinary support
Sense of community—positive attachments to the center by members	No ownership: see center as place for services, no pride in center or sense of belonging	Seem to like each other, but lack communality; not a cohesive group; little ownership or pride in the center	Sense of belonging to the larger group; feel appreciated; sense of ownership and pride
Self-help and reciprocity—sharing information, problem-solving, role-modeling among members	Little or no sharing information or problem-solving among members; helping only when a functional necessity	Few or brief information sharing or problem-solving; share some experiences to get or give advice	Much info exchange, problem-solving, and encouragement toward independence in housing, jobs, and education

TABLE 2: ***Fidelity Ratings for 31 Consumer-Run Drop-In Centers***

Criteria Item	M	SD	*MD*
Structure			
Voluntariness	4.52	1.12	4
Consumer determination of policy, operations, and planning	3.39	1.31	3
Transportation	3.52	1.26	3
Exterior physical environment	3.71	1.42	5
Interior physical environment	3.03	1.14	5
Facilitating referrals	2.87	1.63	5
Outreach to recruit new members	2.74	1.24	3
Activities and services—basic needs	2.81	1.25	5
Housing, transportation, education, and job assistance	1.90	1.25	3
Social recreational activities	2.32	1.33	4
Process—belief systems			
Group empowerment	2.48	1.26	3
Practice or improve social and work-related skills	2.10	1.45	5
Recovery orientation	2.87	1.45	3
Process—opportunity role structure			
Consumer involvement	2.35	1.31	3
Consumer choice and decision-making	4.35	1.31	3
Nonhierarchical structure	3.32	1.40	3
Process—social support			
Member retention	3.00	1.63	5
General respect	4.03	1.25	5
Respect for diversity	3.68	0.87	4
Social support	3.90	1.14	5
Sense of community	4.10	1.01	5
Self-help and reciprocity	3.26	1.77	5

STUDY 2: FIDELITY RATING INSTRUMENT (FRI) REPLICATION STUDY

Sample and Procedures

To test the reliability of the FRI, we conducted site visits at four drop-in centers that were not included in the original sample of 31. We chose programs that reported at least an average of 10 people per day, had been open at least 1 year, and were within 4 hours drive of the Ann Arbor office. The field research director trained two research assistants who were not involved in the original 31 site visits or in constructing the FRI; however, they were familiar with the guiding philosophy behind consumer-centered programs, and they had visited one or two CDIs or similar programs.

The training included appropriate behavior during site visits (including taking observational notes unobtrusively) and using the FRI-CRDI (explanation of the meaning and purpose of each criteria item, possible indicators, and the basis for ratings). Research assistants were told to pay particular attention to physical surroundings, cleanliness, resources available, posted rules and signs, how consumers come and go, conversations between consumers and staff, responses when someone breaks a rule, and staff and consumer interactions. To reinforce the training, the team conducted an all-day, practice site visit at a nearby drop-in center. Each team member rated the program using the FRI-CRDI and afterwards debriefed by discussing their individual rating of each criteria item in detail to clarify and explicate the indicators and ratings.

Following the training, we developed written materials to structure note-taking on relevant observations during the four reliability-testing site visits. The semistructured observation form had a page for each of the FRI criteria containing the criterion and a listing of all the relevant indicators for that criterion (to keep observers aware of what to look for). Observers made notes on this form. The site visit materials also contained an interview to be conducted with the CRDI director to gain information that usually is not observable and of which many consumers at the drop-in may not be aware; for example, whether the center maintains contact with other local community agencies, or how often the center's hours and location have changed. Finally, site visit materials also contained a few additional items for observers to note that could be helpful after the site visit, such as, a count of consumers in attendance by race and gender and observers' estimation of the percentage of consumers whose overt symptoms are frequent or affecting. Later, after the visit, each member of the research team separately used the observation instrument notes and the CRDI director interview to produce independent ratings of the site on each FRI criteria. This activity took approximately 1 hour. The raters then came together to examine their ratings and to discuss areas of disagreement (meetings lasted 1 to 2 hours). The instrument was refined after each site visit to become clearer and more usable, often by adding additional indicators to be observed. Here are some examples of changes made in some of the criteria:

- Interior physical environment: added that furniture should be comfortable and bathrooms clean and contain necessary hygiene supplies.
- Facilitating referrals: expanded the indicators and rating from 3 points to 5, to prioritize the center having knowledgeable staff seeking to link consumers with services rather than focused on having flyers posted and information available on request.
- Self-help and reciprocity: clarified the focus to be about sharing life experiences in engaged supportive conversations, information sharing, helping one another to gain skills, encouraging

one another toward independence in housing, employment, and education.

The training regimen was also improved as deemed necessary. For example, after the first site visit, an additional 1-hour training session was held regarding ways to engage consumers in conversations that are not intrusive but rather helpful to gathering information necessary to rate the program on the fidelity criteria—ways to act friendly and interested and maintain professional boundaries and ways to broach conversation with consumers about their experiences at the center.

Results

The mean fidelity ratings for the four sites in the replication study are presented in Table 3. As is evident, some of the criteria were rated uniformly high across this group of CRDI centers: *accessibility and external safety*, *acceptability and respect* (social environment), and *voluntariness*. Only the latter item was highly rated in the larger sample. It appears that the high ratings on the other two items reflect the fact that the sample for the reliability study was small and was disproportionately from nonurban areas (that is, all the centers located in urban areas had participated in the original study).

Interrater agreement. To analyze interrater agreement on each of the fidelity criteria, we used the methods of generalizability theory (Cronbach, Gleser, Nanda, & Rajarathnam, 1972). These methods examine the relative variance attributable to the object of measurement (e.g., program), to raters or other features of the measurement context, and to the interaction of programs and raters. For variance component estimates, we used the minimum norm quadratic unbiased estimator, a method that makes no distributional assumptions about the data and is appropriate for small samples (Wu, Gumpertz, & Boss, 2001). Although the small sample precluded use of statistical tests, it was possible to estimate for each criterion the proportion of variance attributable to programs, raters, and the rater × program interaction and to calculate generalizability coefficients from these estimates.

Table 4 contains the results of the analysis of generalizability across raters. The generalizability coefficients express the proportion of total variance that can be attributed to systematic differences among programs. These coefficients are analogous to reliability coefficients, with the value *1* indicating perfect interrater agreement and *0* indicating complete lack of agreement. Most of the fidelity criteria showed excellent interrater agreement, with generalizability coefficients of .85 or more. For five criteria, generalizability coefficients could not be computed because the ratings showed no variability across programs on these dimensions; although agreement among raters was nearly perfect, the sampled sites appeared to be so similar on these criteria that they provided insufficient opportunity to assess raters' ability to make distinctions on these programmatic elements. Generalizability coefficients fell below .60 for only three criteria, suggesting that improvement may be needed on conceptualization or specification of criteria for (a) facilitating referrals; (b) housing, transportation, education, and job assistance; and (c) social-recreational activities. Overall, interrater agreement was very good in this small sample, with 18 of the 21 fidelity criteria showing excellent agreement and only three suggesting the need for additional work.

DISCUSSION AND APPLICATIONS TO SOCIAL WORK RESEARCH AND PRACTICE

This report presents a method for operationalizing fidelity criteria into a rating scale, obtaining ratings across diverse programs, and assessing their interrater reliability. Operationalizing the fidelity criteria for CRDI centers was challenging in that these centers developed in communities through grassroots efforts, did not start from a research and demonstration program model, and had no pre-established standards to guide their implementation or operations. The results reported in this article indicate that fidelity criteria for CRDI centers can be operationalized and measured. An examination of the ratings indicated that the 31 centers had great heterogeneity on most of the criteria: All but two of the criteria had ratings across the full range from 1 to 5. Assuming that there is a model for CRDI centers and that following the model reflects the principles of consumerism, this heterogeneity suggests that fidelity criteria and their measurement are needed to ensure quality consumer-run services (CRS).

It is interesting to compare the highly rated criteria in this study to consumer experts' rankings of the importance of the same criteria (Holter et al., 2004). The results show a high level of congruence. Four of the five criteria given the highest ratings by consumerism experts also received the highest ratings for the CRDI centers in our study (*voluntariness*, *acceptability and respect*, *consumer choice and decision-making*, and *sense of community*). Two criteria rated highly by consumerism experts were not among those receiving the top scores in this study: (a) recovery orientation, personal growth and

TABLE 3: ***Mean Fidelity Ratings for the Four Sites in the FRI Replication Study***

Criteria Item	*Mean Rating by Site*			
	1	*2*	*3*	*4*
Structure				
Voluntary	5.00	5.00	4.00	5.00
Consumer determination of policies and operation	5.00	3.67	3.00	1.00
Transportation	3.67	3.00	3.00	1.00
Exterior physical environment	5.00	5.00	5.00	5.00
Interior physical environment	3.00	4.00	3.67	3.00
Facilitating referrals	4.33	2.33	4.33	3.00
Outreach to recruit new members	4.33	3.67	4.33	1.33
Activities and services (provision for basic needs)	1.00	3.67	1.67	3.00
Housing, transportation, education, and job assistance	2.33	1.00	2.33	1.00
Social recreational activities	3.00	2.67	2.00	2.00
Process—belief systems				
Group empowerment	3.67	2.33	3.00	1.00
Practice and improve social and work-related skills	2.33	3.00	3.00	1.00
Recovery orientation, personal growth and development	3.67	3.00	3.00	3.00
Process—opportunity role structure				
Consumer involvement	3.00	3.00	2.33	1.00
Consumer choice and decision-making	5.00	5.00	3.00	5.00
Nonhierarchical structure	4.33	4.33	4.00	2.00
Process—social support				
Member retention	3.00	3.00	3.67	3.00
Social environment—general respect and diversity	4.00	4.00	4.00	4.00
Social support	5.00	3.00	3.00	3.00
Sense of community	4.33	4.33	3.00	2.33
Self-help and reciprocity	4.33	3.00	2.33	2.33

development; and (b) consumer choice and decision-making. Perhaps some of the CRDI centers we rated had consumer members who were highly disabled and less able to make choices on their own behalf, resulting in the lower rating on the latter criteria. Or, perhaps low scores at some centers reflect a hierarchical, rule-dominated environment—just because the staff is made up entirely of consumers does not mean that power differentials will not necessarily develop. The lower score for recovery orientation may reflect the fact that this concept is of rather recent origin and has not yet filtered down to grassroots levels in terms of practices. Or it could be that the CRDI centers studied are weaker than they should be in adopting a recovery and autonomy vision for all members; perhaps, in this sample, for many centers, power and authority resided in a small group, which did not adequately encourage these outcomes across all the membership.

Concerning the interrater reliabilities of the criteria, it was gratifying to find that most of the items had very high generalizability coefficients and that all of the criteria rated as most important by the consumerism experts had coefficients of .89 and above (with the exception of one criteria that failed to have enough variation across the programs studied to enable us to calculate the coefficient). In contrast, none of the three criteria with low generalizability coefficients were rated as important by the consumerism experts (*facilitating referrals*; *housing, transportation, education & job assistance*; and *social-recreational activities*).

Fidelity criteria, such as those presented in this article, are extremely useful for research, evaluation, and quality improvement purposes. Vis-à-vis research, federal funding typically requires that those applying for research funds to determine the efficacy or effectiveness of an innovative program be able to specify its critical ingredients and examine the extent to which fidelity is maintained over time and across sites. The tools to do this involve fidelity rating scales that are valid and reliable, along with program manuals. Thus, having these tools can increase the speed and likelihood of funding for efficacy and effectiveness studies. Such efforts to establish positive outcomes are desperately needed to make CRS a viable model.

Bond, Evans, et al. (2000) described uses of fidelity criteria in multisite research and evaluation studies. That is, across programs, ratings on each fidelity criteria can be related to client outcomes. Fidelity criteria that have significant relationships with client outcomes would be high priority for inclusion in model replications, whereas those unrelated to client outcomes could be seen as

TABLE 4: ***Fidelity Rating Instrument (FRI)—Generalizability Across Raters for 4 Additional Programs, Each Rated by 3 Raters***

	Variance Components			
			Program Generalizability	
Criteria Item	*Program*	*Rater*	*X Rater*	*Coefficient**
Structure				
Voluntary	0.25	0.00	0.00	1.00
Consumer determination of policies and operation	2.67	0.00	0.33	0.96
Transportation	1.22	0.00	0.33	0.92
Exterior physical environment	0.00	0.00	0.00	—
Interior physical environment	0.22	0.00	0.08	0.89
*Facilitating referrals	0.47	0.00	1.58	0.47
Outreach to recruit new members	1.72	0.17	0.92	0.85
Activities and services (provision for basic needs)	1.42	0.00	0.19	0.96
*Housing, transportation, education, and job assistance	0.33	0.00	0.78	0.56
*Social recreational activities	0.14	0.00	0.33	0.56
Process—belief systems				
Group empowerment	1.11	0.11	0.56	0.86
Practice or improve social and work-related skills	0.78	0.00	0.08	0.97
Recovery orientation, personal growth and development	0.00	0.00	0.33	—
Process—opportunity role structure				
Consumer involvement	0.78	0.00	0.33	0.88
Consumer choice and decision-making	1.00	0.00	0.00	1.00
Nonhierarchical structure	1.11	0.00	0.44	0.88
Process—social support				
Member retention	0.00	0.00	0.33	—
Social environment—general respect and diversity	0.00	0.00	0.00	—
Social support	0.00	0.00	0.00	—
Sense of community	0.67	0.00	0.25	0.89
Self-help and reciprocity	0.44	0.00	0.33	0.80

NOTE: Because of the small sample and nonnormal distributions, variance components were estimated via minimum norm quadratic unbiased estimator (Wu, Gumpertz, & Boss, 2001).
*Generalizability below .60. Criteria for which no generalizability coefficient is listed did not vary across programs, so interrater generalizability could not be assessed.

optional. This approach could thus increase the efficiency of programs by eliminating unnecessary elements and improving training of new or replacement staff as to what is really important.

In evaluation and quality improvement studies, measurement of fidelity on rating scales could be a main focus, seeking to relate fidelity to inputs (such as client characteristics, staffing qualifications, funding levels, etc.). Such analyses could help determine necessary resources to run an adequate program as well as the extent to which variations in client characteristics (such as gender and race and ethnicity differences) affect fidelity and may therefore require program adaptations. Fidelity ratings for specific programs or program types could also be examined over time to determine trends and whether changes in program inputs or program activities affected fidelity to the model. Several fidelity researchers (Bond, Williams, et al., 2000; McGrew et al., 1994; Teague et al., 1998) have written about the need to continually assess fidelity to help avoid program drift (unplanned and unsystematic variations away from criteria) in which the less traditional or more difficult to carry out program ingredients are abandoned over time and programs revert to traditional operations. Having quantified measures of fidelity can allow administrators to intervene when program drift is first noticed rather than later when client outcomes are affected.

Our methods appear to be successful in producing reliable fidelity criteria for CRDI centers. Nevertheless, there are limitations to the study that must be acknowledged. Although the data collection was extensive (31 programs and nearly 1000 consumers), it was limited to one state. It could be that in some locations, the activities that reflect the criteria are more difficult to observe, and therefore the fidelity rating scale may be more difficult to apply. Furthermore, although the process outlined appeared to be successful, it was also highly resource-intensive. In the future, when operationalizing the measurement of fidelity criteria for existing program models (vs. those initiated as research and demonstration projects), it will be important to develop a method that is less costly. For example, once the criteria or critical ingredients have been identified, perhaps the benchmarking could be done by experts through structured group decision-making

techniques or with a smaller but highly representative sample of programs.

Other limitations of the approach presented to operationalize and measure fidelity involve the data collection activities of the field staff. The instrument used to document observations, the ISO, could have been improved by being ordered differently and being more structured. That is, the ISO was not organized by criteria and some information that could have later proven useful in rating the programs was not documented on the ISO, although often one or more of the researchers remembered details relevant to the FRI-CRDI indicators and ratings. For example, evidence of efforts that centers made to reach out to the community to recruit new members had not been documented on the ISO. Finally, although we believe the fidelity rating instrument we developed is applicable to many types of CRS, that conclusion needs to be tested.

Having fidelity criteria and methods to rate fidelity to the criteria is increasingly important to social work researchers seeking funding for outcome and effectiveness studies and to social work evaluators and administrators seeking to maintain quality across programs and over time. For those involved with using or developing CRS in mental health, the measure described in this article offers promise as a tool to assess and monitor fidelity to criteria that are integral to CRS. The authors encourage other researchers to use the FRI-CRDI and to collaborate in further expanding the evidence of its reliability and use. (Copies of the full FRI-CRDI measure along with support materials for conducting ratings are available from the lead author.) For those social work researchers involved in program evaluation and program development needing to develop fidelity criteria for new or existing programs, the authors offer the methods described in this article as a template for similar activities in other programmatic areas.

REFERENCES

Blakely, C. H., Mayer, J. P., Gottschalk, R. G., Schmitt, N., Davidson, W. S., Roitman, D. B., et al. (1987). The fidelity-adaptation debate: Implications for the implementation of public sector social programs. *American Journal of Community Psychology*, *15*, 253-268.

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., Evans, L., Salyers, M. P., Williams, J., & Kim, H. K. (2000). Measurement of fidelity in psychiatric rehabilitation. *Mental Health Services Research*, *2*, 75-87.

Bond, G. R., Williams, J., & Evans, L. (2000). *PN-44-Psychiatric rehabilitation fidelity toolkit*. Cambridge, MA: Human Services Research Institute.

Borkman, T. (1990). Self-help groups at the turning point: Emerging egalitarian alliances with the formal health care system? *American Journal of Community Psychology*, *18*, 321-332.

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.

Chamberlin, J. (1984). Speaking for ourselves: An overview of the ex-psychiatric inmates' movement. *Psychosocial Rehabilitation Journal*, *8*, 56-63.

Chamberlin, J., & Rogers, J. A. (1990). Planning a community-based mental health system: Perspective of service recipients. *American Psychologist*, *45*, 1241-1244.

Chamberlin, J., Rogers, E. S., & Ellison, M. (1995). Self-help programs: A description of their characteristics and their members. *Psychiatric Rehabilitation Journal*, *19*, 33-41.

Connnelly, L. M., Keels, B. S., Kleinback, S. V., Schneider, J. K., & Cobb, A. K. (1993). A place to be yourself: Empowerment from the client's perspective. *Image Journal Nursing School*, *25*, 297-303.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajarathnam, N. (1972). *The dependability of behavioral measures: Theory of generalizability for scores and profiles*. New York: Wiley.

Davidson, L., Chinman, M., & Kloos, B. (1999). Peer support among individuals with severe mental illness: A review of the evidence. *Clinical Psychology: Science & Practice*, *6*, 165-187.

Donabedian, A. (1980). *The definition of quality and approaches to its assessment*. Ann Arbor, MI: Health Administration.

Hohmann, A. A., & Shear, K. M. (2002). Community-based intervention research: Coping with the "noise" of real life in study design. *American Journal of Psychiatry*, *159*, 201-207.

Holter, M. C., Mowbray, C. T., Bellamy, C., MacFarlane, P., & Dukarski, J. (2004). Critical ingredients of consumer run services: Results of a national survey. *Community Mental Health Journal*, *40*, 47-63.

Kennedy, M., Humphreys, K., & Borkman, T. (1994). The naturalistic paradigm as an approach to research with mutual-help groups. In T. J. Powell (Ed.), *Understanding the self-help organization: Frameworks and findings* (pp. 172-189). Thousand Oaks, CA: Sage.

Macias, C., Propst, R., Rodican, C., & Boyd, J. (2001). Strategic planning for ICCD clubhouse implementation: Development of the Clubhouse Research and Evaluation Screening Survey (CRESS). *Mental Health Services Research*, *3*, 155-167.

Maton, K. I., & Salem, D. A. (1995). Organizational characteristics of empowering community settings: A multiple case study approach. *American Journal of Community Psychology*, *23*, 631-656.

McGrew, J. H., Bond, G. R., Dietzen, L., & Salyers, M. (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). Fidelity to assertive community treatment and client outcomes in New Hampshire dual disorders study. *Psychiatric Services*, *50*, 818-824.

Mowbray, C. T. (1997). Benefits and issues created by consumer role innovation in psychiatric rehabilitation. In C. T. Mowbray, D. P. Moxley, C. Jasper, & L. L. Howell (Eds.), *Consumers as providers in psychiatric rehabilitation* (pp. 45-63). Columbia, MD: International Association for Psychosocial Rehabilitation Services.

Mowbray, C. T., Moxley, D. P., Jasper, C., & Howell, L. L. (1997). *Consumers as providers in psychiatric rehabilitation*. Columbia,

MD: International Association for Psychosocial Rehabilitation Services.

Mowbray, C. T., Robinson, E. L., & Holter, M. C. (2002). Consumer drop-in centers: Operations, services, and member involvement. *Health & Social Work*, *27*, 248-261.

Orwin, R. G. (2000). Assessing program fidelity in substance abuse health services research. *Addiction*, *95*, S309-S327.

Paulson, R. I., Post, R. L., & Herinckx, H. A. (2002). Beyond components: Using fidelity scales to measure and assure choice in program implementation and quality assurance. *Community Mental Health Journal*, *38*, 119-128.

Salem, D. A. (1990). Community-based services and resources: The significance of choice and diversity. *American Journal of Community Psychology*, *18*, 909-915.

Segal, S. P., Silverman, C., & Temkin, T. (1993). Empowerment and self-help agency practice for people with mental disabilities. *Social Work*, *38*, 705-712.

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

Test, M. A. (1998). Community-based treatment models for adults with severe and persistent mental illness. In J. B. W. Williams, & K. Ell (Eds.), *Advances in mental health research: Implications for practice* (pp. 420-436). Silver Springs, MD: National Association of Social Workers.

Von Tosh, L., & Delvecchio, P. (1998). *Consumer/survivor-operated self-help programs: A technical report*. Rockville, MD: Center for Mental Health Services, Community Support Branch.

Wu, C., Gumpertz, & Boss, D. D. (2001). Comparison of GEE, MINQUE, MI, and REML estimating equations for normally distributed data. *The American Statistician*, *55*, 125-130.

Yanos, P. T., Primavera, L. H., & Knight, E. L. (2001). Consumer-run service participation, recovery of social functioning, and the mediating role of psychological factors. *Psychiatric Services*, *52*, 493-500.