Regular Article

Is It Wraparound Yet? Setting Quality Standards for Implementation of the Wraparound Process

Eric J. Bruns, PhD
Jesse C. Suter, PhD
Kristen Leverentz-Brady, PhD

Abstract

The wraparound process has increasingly been used as a mechanism to plan and coordinate services for children with behavioral health needs and their families. This has led to growing interest in assessing wraparound implementation against standards for quality. However, there has been little consideration of how best to establish such benchmarks or guidelines. Using both a norm-referenced and criterion-referenced approach, this study established preliminary criteria for assessing the adequacy of wraparound implementation using the Wraparound Fidelity Index, version 3, a multi-informant interview that assesses conformance to wraparound principles. The evaluation system was then applied to ten wraparound programs and 11 different study samples assessed via the Wraparound Fidelity Index version 3 in research studies. The system was constructed to discriminate different wraparound conditions assessed in research studies while still being attainable by the ten established wraparound programs. Implications for evaluating wraparound programs and for setting fidelity benchmarks in behavioral health services research are discussed.

As interest grows in implementing effective behavioral health interventions, so does interest in assessing implementation against established standards of quality. However, the need for instruments that can provide such information has tended to outstrip the science of creating, validating, and applying them.[1]

Once fidelity instruments are created, a particular challenge is presented by the practical question of knowing when adherence scores are "good enough." Such challenges are especially vexing for service

Addess correspondence to Eric J. Bruns, PhD, Department of Psychiatry and Behavioral Sciences, Division of Public Behavioral Health and Justice Policy, University of Washington School of Medicine, 2815 Eastlake Ave E, Suite 200, Seattle, WA 98102, USA. Phone: +1-206-6852477. Fax: +1-206-6853430. Email: ebruns@u.washington.edu

Jesse C. Suter, PhD, Center on Disability and Community Inclusion, University of Vermont, 311 Mann Hall, 208 Colchester Ave., Burlington, VT, 05405-1757 USA.

Kristen Leverentz-Brady, PhD, Department of Psychology, University of Vermont, John Dewey Hall Burlington, VT, 05401 USA.

Journal of Behavioral Health Services & Research, 2008.

delivery models that are individualized, not strictly manualized, and/or represent service processes rather than regimented treatment procedures.[2] When a treatment process is complex or individualized or does not have a protocol for what exactly is supposed to happen in treatment, establishing benchmarks for use in quality assurance and policy decisions will be more difficult than assessing whether the clinician(s) did the prescribed activities. Instead, assessment of the adequacy of adherence may require an empirical process based on available data from multiple sites or programs.

Salyers and colleagues tackled this difficult issue in attempting to "bootstrap" standards for Assertive Community Treatment (ACT), a team treatment approach characterized by frequent team meetings and the development of plans of care for adults with severe mental illness that emphasize community-based services and daily living skills.[1,3] Their method was guided by two relevant options for interpreting a site or community's fidelity assessment results. The first was a *norm-referenced* approach, whereby mean scores for one program site were compared to a large group of program sites. The second method followed a *criterion-referenced* approach, in which a score was compared to an external standard, such as one related to prediction of performance (e.g., a score that has been found to be associated with positive client outcomes or better service delivery). In the current paper, a similar process is described for the wraparound process for children, youth, and families.

The Wraparound Process

Like ACT, wraparound is a treatment *process* rather than a single intervention for a specific problem area. Also like ACT, its goal is to ensure that the individual is adequately supported to be maintained in the community. However, in the case of wraparound, the focus of the effort is typically children with serious emotional and behavioral problems and their families.[4] The wraparound research base is still "on the weak side" with respect to utilization of rigorous designs.[5] Nonetheless, results of evaluation studies,[4,6,7] success stories from individual communities,[8–10] and clear popularity with communities and families[11,12] have led to wraparound's inclusion in two Surgeon General's reports,[13,14] the recommendation for its use in federal grant programs,[15] and the discussion in the research literature of its integration with evidence-based practices.[12,16,17]

Similar to ACT, increased visibility and implementation in an era of greater use of data in decision making has led to questions about how state authorities and local collaborative bodies can best make policy and funding decisions around adequacy of wraparound implementation. Though many wraparound trainers and local service systems have established their own approaches to measuring wraparound quality and fidelity,[2] to date, the most commonly used instrument for assessing wraparound implementation is the Wraparound Fidelity Index (WFI).

The Wraparound Fidelity Index The WFI assesses adherence to the established principles of wraparound for an individual youth and his or her family through interviews with multiple respondents. The measure has demonstrated test–retest and interrater reliability and validity in the form of association with external expert ratings.[2] The WFI has also demonstrated discriminant validity in the form of significant differences for wraparound vs. non-wraparound comparison sites and sites with different levels of implementation support.[18] WFI scores have also been found to associate with child and family outcomes.[10,19] However, data from these national studies have not yet addressed the question of what fidelity scores show that wraparound is "good enough."

The Current Study

Given the goal of assisting jurisdictions that are implementing wraparound to better interpret results of their fidelity assessments, a process was undertaken similar to that completed by Salyers

and colleagues around fidelity standards for ACT.[1] In the current study, a combination of norm referencing and criterion referencing was employed to develop fidelity benchmarks for the WFI-3 using available data from a national multi-site dataset, as well as evaluation studies that used the WFI-3. Our goal was to make preliminary recommendations about thresholds for fidelity to the wraparound principles as assessed by the WFI-3. At the same time, communities and programs who are truly implementing wraparound should reasonably be expected to achieve any such benchmarks.

To accomplish the above goals, a three-step approach was employed. First, the national WFI-3 dataset was examined to determine typical scores for a sample of wraparound programs. Second, studies were compiled and reviewed that collected WFI-3 data as well as data for one or more external criteria variables. Two types of external criteria were sought: child and family outcomes (which should be predicted by fidelity scores) or degree of implementation supports for wraparound (which should predict fidelity scores). Third, results of steps one and two were used to propose preliminary fidelity standards and assess the number of sites in the national sample that attained these standards. The overall goal was to use the results of these three exercises together to identify reasonable and attainable wraparound fidelity benchmarks for a community or program.

Method

Measure

Wraparound adherence The WFI is a structured interview that assesses adherence to 11 core principles of wraparound (family voice and choice, team-driven, individualized, natural supports, community-based, culturally competent, strengths-based, unconditional care, collaboration, flexible resources, outcomes-based). Four items serve as indicators for each principle, with responses ranging from 0 (low fidelity) to 2 (high fidelity). The Resource Facilitator (RF) form (usually completed by a service provider and/or care coordinator) and Caregiver (CG) form are parallel to one another, each containing 44 similar items (with differences in wording that is specific to each respondent). There is also a Youth (Y) form that includes 32 items. Examples of items (taken from the Caregiver form) are presented below:

- As the parent or primary caregiver, are you given highest priority when making major decisions? (Principle 1: Family Voice and Choice)
- Is there a friend or advocate of your family who actively participates on the team? (Principle 2: Youth and Family Team)
- Does the team help your child get involved with activities in your community? (Principle 3: Community-Based Services and Supports)
- Does the team help you receive support from your friends and family? (Principle 7: Natural Supports)

Scores are summed to create total scores, which in turn are converted to percentages (0% to 100%) to ensure consistency of scale scores across respondents and for ease of interpretation. For the present study, a single overall score for each family was calculated by averaging the separate respondent scores (resource facilitator, caregiver, youth).

Using data from a national WFI-3 sample (employed in the current study) of $N=667$ youth served by ten wraparound initiatives in nine states, internal consistency for all items of the WFI-3 (i.e., total fidelity scores) has been found to be strong, with Cronbach alpha coefficients equaling 0.82 for the Wraparound Facilitator form, 0.91 for the Caregiver form, and 0.84 for the Youth form. Using the same national sample, interrespondent agreement for the WFI-3 was assessed by calculating the intraclass correlation coefficient (ICC) for agreement between sets of respondents.

Results found ICCs of 0.58 for all three respondents, 0.44 for RF–CG agreement, 0.49 for CG–Y agreement, and 0.45 for RF–Y agreement. These ICCs indicate moderate to strong interrespondent agreement for a scale of this nature.

Two-week test–retest reliability of the WFI was established in a study involving two wraparound programs in two states. Results found reliability to be moderate to strong, with Pearson r correlations found to be $r = 0.84$ for the RF form, $r = 0.88$ for the CG form, and $r = 0.64$ for the Youth form.[2] Finally, as described above, evidence for construct validity is found in studies demonstrating significantly different total WFI scores for samples of wraparound vs. non-wraparound-enrolled youth.[19,20] Evidence for criterion-related validity is provided by studies that have found significant associations between total WFI scores and child and family outcomes.[10,21]

For the current study, the overall fidelity score, determined by averaging scores from the three respondents, was used because of its utility as an overall measure of fidelity, strong reliability in the above psychometric studies, and ability to discriminate sites in past validity studies.[2,22]

National WFI sample

This sample was derived from a validation trial of the WFI.[2] Agencies or sites providing wraparound were recruited though conference presentations and formal and informal networking. Those that were interested in using the WFI completed a formal request, which was reviewed by the research team and accepted if proposals showed evidence that the site was (1) implementing wraparound, (2) planned to recruit a representative sample of participating families, and (3) agreed to administer the WFI as stated in the manual, by trained interviewers who were not involved in service delivery with the families being assessed.

Once approved, sites in the national sample recruited families to participate. Specific guidelines were provided to participating sites about sample construction. For example, sites were directed to recruit families of youth who had been enrolled in wraparound for an adequate duration (e.g., at least 3 months) to allow enough experience with the service process to be able to respond to fidelity items. Participating sites differed in their sampling methods and approaches to recruiting youth and families but were provided with guidance by the research team. For example, larger sites were encouraged to conduct random stratified sampling, while smaller sites often attempted to conduct WFI interviews with families and providers of all enrolled youth.

Local evaluation teams obtained informed consent using a local adaptation of a boilerplate consent form provided by the research team. A lead local evaluator (who met criteria for leading local implementation of the WFI protocol and who maintained contact with the WFI research team) trained interviewers based on the WFI user's manual.[22] The local evaluators then administered interviews, entered data into a database provided by the research team, and forwarded data (with identifying information deleted) to the research team. Response rate or rate of completion of interviews for participating sites is not known.

National WFI sample sites The current study included WFI-3 data for ten sites with data across the three WFI-3 respondents for at least $n = 20$ enrolled youth. The ten sites were located in nine US states, had been in operation from 2.5 to 7 years ($M = 3.9$, SD = 1.45), and had served between 41 and 497 families during their existence ($M = 223$, SD = 164). Such wide variation in numbers of youth and families served is typical of wraparound projects nationally, which may be administered by small service agencies or across large jurisdictions.[23]

Two sites in the national sample were primarily aimed at youths in state custody, six programs explicitly served youth "eligible for/at risk of placement in high-level residential care," and two programs were specifically tailored to serve youths transitioning from group homes or residential treatment facilities. The number of families on which WFI-3 data were available ranged from

20–370 across sites (M= 66.7, SD=107.6; the wide range and large SD in the number of families is primarily accounted for by one large county-based initiative that had been collecting data for several years). For sites that assessed fidelity for individual youth at multiple points over time, only the first assessment was included in the national sample used in this study.

Characteristics of national WFI sample Fidelity assessment data for a total of N=667 youths were included in the national sample across the ten sites. At the time of the WFI administration, youths/families had been receiving wraparound for a mean of 13.4 months (SD=9.04), with a range of 1 to 88 months. The relatively large mean and SD reflected in these data is somewhat accounted for by the presence of two wraparound sites that served youth in state custody, many of whom received support through the wraparound process during long stays in foster care.

The majority of the youths were male (68%), and ages ranged from 2 to 21 years (M=13.8, SD=3.27). Three hundred ninety-three (59%) of the youth were identified as Caucasian, 65 (10%) as African American, four as Native American, four as Asian/Pacific Islander, 14 as biracial, and 34 (5%) as "other race" (race was not identified for the remaining 157 youths or 23%). Nineteen youth (3%) were identified as Hispanic or Latino. At the time of the WFI interviews, the majority of youth (51%) resided with one or more biological (48%), adoptive (2%), or stepparents (1%); 10% resided with other relatives (7% grandparents); 15% resided with foster parents; and 15% were in other types of residential settings, including group homes, residential treatment, and psychiatric hospitals (residential placement was not reported for 9% of the sample). Three hundred four (46%) of the youths in the sample were in state custody at the time of WFI interviews.

Procedure

Norm referencing Mean WFI-3 overall fidelity scores were examined at the site level for the ten programs included in the study (total N=667 families; see Table 1). Analyses of variance (ANOVA) with Tukey's post hoc analyses were conducted to determine significant differences across these sites.

Criterion referencing A review was conducted of studies published, in press, or presented at a major conference that presented WFI-3 data for two or more groups that differed with respect to

Table 1
Overall mean WFI-3 scores for the ten national study sites

Site	Number of participants	*M* (SD)
1	74	72.2 (11.0)
2	22	72.4 (9.4)
3	31	73.5 (7.2)
4	21	75.1 (8.9)
5	27	75.3 (11.7)
6	26	79.4 (8.9)
7	20	79.5 (10.0)
8	370	79.7 (10.8)
9	34	80.0 (6.6)
10	49	80.1 (8.9)
Site-level mean		76.7 (2.3)

WFI-3 Wraparound Fidelity Index, version 3

one or more external criteria. "External criteria" could include (1) data on outcomes for two or more independent groups, (2) specific information describing differences in the state of development of or supports for the wraparound program (e.g., just starting services vs. many years of implementation, greater resources for implementation vs. fewer resources, presence of training vs. no training), or (3) specific information describing differences in the nature of services delivered (e.g., a study of wraparound vs. a comparison or treatment as usual condition).

Eligible studies for this review were identified through electronic and manually based searches of literature published through January 31, 2006. Electronic searches were conducted via the Web of Science, PsycINFO, and ERIC electronic databases using the keywords: *wraparound* and *Wraparound Fidelity Index*. In addition, a manual search was conducted of the annual research conference proceedings of *A System of Care for Children's Mental Health: Expanding the Research Base* hosted by the University of South Florida, Louis de la Parte Florida Mental Health Institute, Research and Training Center for Children's Mental Health. This source was chosen because it has been a primary outlet for research on wraparound. The review included some unpublished and non-peer-reviewed sources (e.g., conference proceedings) due to the small total number of studies. However, data from annual reports, newsletters, or other non-research outlets were not included in this review.

Five studies were included. Two were evaluation studies that included WFI-3 scores for a wraparound service delivery condition vs. a non-wraparound comparison or control condition.[20,24] One study presented WFI-3 data for a sample of programs for which the system and organizational supports for wraparound had been measured by a separate instrument, and for which each program had been characterized as having relatively higher vs. lower scores on this measure (by median split).[25] A fourth study presented WFI-3 data for a group of wraparound facilitators who served youth that demonstrated significant improvement in functioning as assessed by the Child and Adolescent Functional Assessment Scale (CAFAS)[26] vs. a group of facilitators who served youth that did not show improvements in functioning.[10,27] The fifth study presented WFI-3 data for three stages of a program, across which wraparound implementation support increased at each stage. Specifically, in this study, wraparound implementation was assessed via the WFI-3 before the initiation of provider training on wraparound implementation, after provider training but before initiation of provider coaching and after provider training and coaching.[28] A summary of the five studies included in the criterion-referencing exercise is presented in Table 2.

Fidelity benchmarks and attainability analysis Based on the distribution of WFI-3 overall scores from sites in the national sample and observed differences in mean scores for groups with different external criteria, the authors proposed preliminary fidelity benchmarks. These preliminary benchmarks were then tested against WFI-3 Overall Fidelity scores from the sites in the national sample as well as the different conditions included in the studies reviewed. The number and percent of the ten national sites and of the 11 cohorts from the studies reviewed were identified that achieved each of the different proposed benchmarks for wraparound fidelity. The authors then adjusted the proposed benchmarks to improve interpretability (more details on this process and its challenges are presented in "Results" and "Discussion" below.)

Results

Norm referencing

Table 1 presents overall fidelity scores for the ten sites included in the norm-referencing exercise in ascending order of mean scores obtained by each site. As shown, the mean overall fidelity scores for the ten sites ranged from 72.2% to 80.1%, with a site-level mean of 76.7% (SD=2.3). Results of one-way ANOVA found significant differences across all sites [$F(9, 656)=5.951$; $p<0.0001$],

Table 2
Results of studies employing the WFI-3 included in criterion-referencing exercise

Study	Study groups, *n* youth, type of study	WFI-3 scores from study results
Studies of different treatment conditions		
Peterson et al.[24]	Wraparound (*n*=33) vs. treatment as usual (*n*=32); matched comparison study	75% for wraparound group; 60% for comparison group
Ferguson[20]	Wraparound (*n*=49) vs. treatment as usual (*n*=30); randomized control study	75% for wraparound group; 63% for control group
Studies of different external criteria		
Rast et al.[27]	Youth who exhibited more positive (*n*=8) vs. less positive (*n*=24) outcomes; cross-sectional study	87% for more positive outcomes; 72% for less positive outcomes
Bruns et al.[18]	Wraparound with higher organizational and system support scores for implementing wraparound (2 sites; total *n*=44) versus lower support scores (2 sites; total *n*=67); cross-sectional study	84% for two programs with greater support; 72% for programs with lesser support
Studies of changing training conditions over time		
Rast and VanDenBerg[28]	Wraparound implementation pre-training (*n*=21), post-training (*n*=32), and post-training + coaching (*n*=30); single-site longitudinal study	64% pre-training, 72% post-training, 86% post-training and coaching

WFI-3 Wraparound Fidelity Index, version 3

with results of Tukey post hoc analyses indicating that sites 1, 2, and 3 (overall fidelity≤73.5%) were significantly different from sites 8, 9, and 10 (≥79.7%). Site 8 was also found to be significantly different from sites 4 and 5 likely due to its much larger sample size.

Criterion referencing

Table 2 presents a summary of the review of studies employing the WFI-3 that included different treatment conditions or that were found to differ on some type of external criteria. With respect to different treatment conditions, two studies were found that assessed fidelity for both wraparound and non-wraparound comparison groups.[20,24] These studies found significant between-group differences in WFI-3 scores, with WFI-3 overall fidelity scores of 60% and 64% for non-wraparound groups compared to 75% and 76% for wraparound groups, respectively.

Two studies were found that presented WFI-3 scores for samples or sites with differing characteristics. A study of WFI-3 scores for sites with measurably fewer versus greater organizational and system supports for wraparound (as assessed by a structured interview) found that sites with fewer supports achieved mean WFI-3 scores of 72% compared with 84% for sites with more documented supports for wraparound.[18] A second study examined WFI-3 scores for individual wraparound facilitators whose clients demonstrated greater improvements on the CAFAS vs. facilitators whose clients showed less improvement in functioning. Facilitators with clients who experienced greater

improvements in functioning were found to have mean WFI-3 scores of 87%, compared to 72% for facilitators whose clients experienced less improvement.[27]

Finally, a longitudinal study was found of one site in which providers received different levels of training and coaching over time. Results from this study found that mean WFI-3 scores increased from 64% pre-training to 72% post-training, to 86% after implementation of both training and coaching.[28]

Determination of Benchmarks

In a final stage of analysis, a straightforward and useful system for assigning guidance to communities and programs in interpreting overall WFI-3 fidelity scores was attempted. Criteria used to evaluate the proposed grading system were that, first, it clearly differentiate between the different study samples (e.g., wraparound and non-wraparound programs; more-positive vs. less-positive outcomes; greater vs. fewer implementation supports); and second, attainability (i.e., that the wraparound programs in the national sample would be able to attain "passing" scores).

After reviewing the results of studies reviewed and the range of site-level WFI scores found from the national dataset, we aimed to construct a system of benchmarks that would assign "failing" grades to non-wraparound sites, high marks to those achieving superior outcomes or with greater supports for wraparound, and a range of grades in between. After some trial and error, a system was chosen that characterized sites or programs scoring at 85% or higher as "high fidelity," based on the observation that facilitators whose clients experienced the most positive outcomes, and sites with the greatest level of organizational and system supports for wraparound implementation scored above 85%. Additionally, it was clear that overall WFI-3 scores above 90% were very difficult to obtain, as such a score was not found for any sites or study cohorts.

Below 85%, five-point increments were used for ease of use and interpretation. In this system, scores of 80–84% were characterized as above average, 75–79% as average (reflecting that the mean site-level score from the national dataset was 77%), and 70–74% as below average. Sites or programs below 70% were characterized as "non-wraparound or in need of substantial improvement," based on the observation that scores below 70% were only found for non-wraparound comparison sites and sites just beginning implementation of wraparound, before provider training.

Table 3 presents the results of applying this system to all the sites included in the norm-referencing and criteria-referencing exercises. As shown, this system discriminated sites that represented different conditions. All non-wraparound conditions fell in the lowest category (non-wraparound/needing substantial improvement). All conditions representing more-positive outcomes or higher system support scores were characterized as "high fidelity" or "above average." The five sites and conditions representing "standard wraparound conditions" (i.e., wraparound cohorts in program evaluations, wraparound facilitators achieving less positive outcomes, or sites with fewer supports) fell into the "average" or "below-average" range. This system also ensured attainability for sites in the national WFI-3 sample, with 100% of sites "passing." At the same time, there was a range of grades achieved across the sites in the national sample: Two of the ten sites fell in the "above-average" range, 50% in the "average" range, and 30% in the "below-average" range.

Discussion

The wraparound process is increasingly being adopted as a mechanism for communities to achieve more-positive outcomes for youth with behavioral and emotional challenges, with the model being written into lawsuit settlement implementation plans, state legislation, and other state and local policy documents.[29,30] As this has occurred, our research team has received increasing inquiries about how the WFI can be used to evaluate the achievement of "high-fidelity" wraparound that has been found to be associated with improved outcomes for youth in many evaluations.[8–10] In this study, a

Table 3

Percent of study sites and study conditions achieving wraparound fidelity report card cutoff scores

	n (%)				
Program/study condition	**High fidelity (85–100%)**	**Above average (80–84%)**	**Average fidelity (75–79%)**	**Below average (70–74%)**	**Non-wraparound/ needing substantial improvement (0–69%)**
National sample sites (*N*=10)	0	2 (20%)	5 (50%)	3 (30%)	0
Non-wraparound and "pre-training" wraparound conditions (*N*=3)	0	0	0	0	3 (100%)
Standard wraparound conditions (*N*=5)	0	0	2 (40%)	3 (60%)	0
More supports/more positive outcome wraparound conditions (*N*=3)	2 (67%)	1 (33%)	0	0	0

WFI-3 Wraparound Fidelity Index, version 3

preliminary system for interpreting WFI-3 overall fidelity scores based on norm and criterion referenced standards was developed. Central to this exercise was the presence of a number (albeit small) of control and comparison studies, and studies that link fidelity scores to external criteria.

Based on data from a national sample and a review of studies that have used the WFI-3, representing 21 sites or study conditions overall, WFI-3 scores were found to range from 60% to 87% of the total possible score. A pattern was identified whereby non-wraparound groups scored in the low 60s and wraparound with greater system supports for implementation or association with more-positive outcomes scored in the mid- to high 80s. Setting descriptive benchmarks based on these findings yielded "grades" for sites in the national sample that fell into a bell curve across grades of "below average" to "above average" with all sites achieving "passing" grades but no sites achieving a score over 80%. Thus, none of the sites in the national sample achieved a grade of "high fidelity," which would require a site-level score of 85% or higher.

This seemed reasonable with respect to both attainability and our goal of setting high standards for sites based on evidence for what levels of fidelity represent potential for greater positive impact. This grading system and its rationale has been included in an updated version of the WFI-3 user's manual[22] and in guidance to researchers and program administrators using the WFI-3. It is worth pointing out that, initially, the set of ranges in the system described above were described using traditional academic letter grades from A (for sites scoring ≥85%) to F (for sites scoring <70%). However, because most academic scoring systems would not assign an "A" for a score of 87% (the highest score obtained among groups in our set of reviewed studies), it was decided to abandon academic letter grades in favor of the one described above, which is based more straightforwardly on the distribution of scores found in this study.

Unlike the exercise in developing standards for ACT conducted by Salyers and colleagues,[1] the current study did not conduct an item-level analysis to determine which items best discriminated sites or conditions on different criteria. This is because previous studies have shown the Overall

Fidelity score to have the best criterion and construct validity[22] and also because the WFI's construction as a set of linked interviews has led to a convention of using overall fidelity scores to evaluate overall fidelity. Nonetheless, this may represent a shortcoming of the proposed approach to setting standards. First, certain items are clearly more useful in evaluating areas of particular challenge for sites. Previous research using the WFI has shown that programs often have difficulty achieving diverse wraparound teams, balancing formal and informal strategies in wraparound plans, having adequate flexible funds to implement creative strategies, and achieving other criteria for high-quality implementation.[25] Items that assess these areas may provide information that is more useful and discriminatory than others for which there is less variability.

This problem is also demonstrated by data presented in Tables 1 and 2, whereby ten national programs differ only by eight percentage points, and non-wraparound sites still achieve total scores of over 60%. This restricted range of scores is likely an artifact of the interview method employed by the WFI-3, whereby social desirability demands on the part of family members and concerns about performance reviews on the part of staff may result in more positive scores on the more subjective items. Overall, the use of interviews is viewed as a positive component of wraparound fidelity measurement because it involves participants and providers in a quality improvement process that helps educate about and reinforce the principles of wraparound. Nonetheless, future versions of the WFI are now being piloted that include data from additional sources, such as team meeting observation and document reviews.

In addition to the restricted range of scores, it is notable that no sites were found to score over 87%. This is due to several WFI items consistently scoring very low, such as whether the team relies primarily on professional services rather than a mix of natural and formal supports, whether school or other key agency representatives had a role on the team, and whether the wraparound team succeeded in getting the youth involved in community activities. Though inclusion of items that are consistently "difficult" may further restrict the variance of scores, the principles and specified activities of the wraparound model promote such implementation components as ideals for communities and care management organizations to strive for. It is expected that continuing to include such items will help promote development of supports to achieve fidelity in these areas in service systems implementing services via the wraparound model.

Study limitations

In addition to limitations of the WFI-3, limitations of the current study need to be acknowledged. First, the number of useful studies found in the review was small. If discernable patterns were not so clearly observable across the 11 conditions presented in this set of studies, results of the criteria-related exercise would likely have been inconclusive. Moreover, several assumptions are being made about the rigor of the research employed in the studies used in the criteria-referencing exercise. Three of the studies reviewed were presented at conferences rather than published in peer-reviewed journals, and data collection methods varied across these sites. For these reasons, we employed the norm-referencing exercise sites from our national sample in addition to the criterion-referencing exercise. Nonetheless, the conclusions drawn and grading system derived from this combination of approaches should be considered preliminary guidance more than a strict set of standards.

Finally, we are making an additional assumption about the nature of the wraparound programs in our national WFI-3 sample. Although programs included in this sample were screened before being accepted and agreed to adhere to data collection standards, it should be recognized that this was an opportunity sample, and WFI-3 data collection methods varied across these sites. Differences in the sites or programs themselves are also not well understood and may be a result of many confounds (e.g., geographic characteristics, client samples, state regulations). Continuing development of

wraparound fidelity scales and standards will necessarily depend on similar "real-world" opportunities but will ideally also be based on more systematic studies that include consistent data collection and links to well-understood external criteria. Several such studies are now underway.

Implications for Behavioral Health

The WFI has become a widely used instrument for several reasons, including the widespread implementation of wraparound and related models, the measure's ease of use, and its direct involvement of family members and providers, potentially increasing opportunities for reinforcing practice principles. Results can also be applied to quality improvement activities, supervision, and policy development. At the same time, these practical utilities have not aided interpretation of WFI results.

To aid in this application of the WFI, a benchmarking exercise was undertaken, and a preliminary guide to interpreting WFI scores was developed. Study and scale limitations prevent assertions that the resulting guidance should be used for purposes such as certification or accreditation. Nonetheless, the results should be useful to those aiming to interpret their local WFI-3 overall fidelity and respondent total scores. Results also present another example of methods that may be used by services researchers to set criteria for quality using fidelity scale scores.

Results of the exercise will also prove useful in interpreting research results. For example, a recent study[31] compared outcomes for a large group of youth with serious emotional disturbance whose services were coordinated via the wraparound process to a small cohort of youth in the same system of care who were selected to receive multisystemic therapy (MST), an intensive, evidence-based intervention for youth with serious behavior problems.[32] Both groups showed significant improvements in functioning as assessed by the CAFAS and behavior as assessed by the Child Behavior Checklist, and similar downward trajectories in scores for each of these measures.[33] Rates of improvement in child functioning over time were not significantly different for the two groups; however, improvement in behavior problems was significantly better for the MST group.

The study reported mean overall adherence scores from the WFI-3 of 82–83% over the course of the study (MST adherence data were not available). Using the guidelines developed in the current study, such WFI scores would be characterized as "above average." In addition, it is clear that model adherence in this system of care was quite high compared to sites nationally. Such information helps us interpret the level of wraparound fidelity that may be necessary for a program or community to demonstrate functional and behavioral outcomes comparable to those achieved in this system of care.

Future research

The current study and other activities involving wraparound fidelity measurement have raised awareness of the potential utility and interpretability of instruments such as the WFI while also pointing to directions for future research and development. For example, future iterations of the WFI should include methods to train interviewers to criteria. Fidelity measurement systems should also incorporate sources of data beyond interviews, as well as better measures of the organizational and system context within which wraparound implementation is occurring. Efforts to better specify wraparound as a care management model are aiding this process.[29,30]

In the short term, however, there is an immediate need for practical methods for measuring the quality of implementation of the wraparound process and benchmarks for interpreting scores. Both research[2,10,21] and theory[4,12,17,34] have generated the expectation that high-quality wraparound practice will improve outcomes for children and youth. At the same time, collection and use of fidelity and outcomes data is considered fundamental to successful implementation of any service

delivery model.[35,36] Within children's behavioral health services, robust adherence to the principles of wraparound is likely to represent a fundamental level of quality to which all communities should strive as it supports these youth and their families.

Acknowledgment

This research received support from several sources, including the Center for Mental Health Services, Substance Abuse and Mental Health Services Administration; the Child, Adolescent, and Family Unit of the Vermont Department of Developmental and Mental Health Services; and ORC Macro, Inc. We would like to thank April Sather, MPH, for her assistance in manuscript preparation.

References

1. Salyers MP, Bond GR, Teague GB, et al. Is it ACT yet? Real-world examples of evaluating the degree of implementation for assertive community treatment. *Journal of Behavioral Health Services & Research*. 2003;13:304–320.
2. Bruns EJ, Burchard JD, Suter JC, et al. Assessing fidelity to a community-based treatment for youth: the Wraparound Fidelity Index. *Journal of Emotional and Behavioral Disorders*. 2004;12:79–89.
3. Test MA, Stein LI. Community treatment of the chronic patient: Research overview. *Schizophrenia Bulletin*. 1980;4:350–364.
4. Burchard JD, Bruns EJ, Burchard SN. The wraparound process. In: Burns B, Hoagwood K, English M, eds. *Community-Based Interventions for Youth*. New York, NY: Oxford University Press; 2002:69–90.
5. Farmer EMZ, Dorsey S, Mustillo SA. Intensive home and community interventions. *Child and Adolescent Psychiatric Clinics of North America*. 2004;13:857–884.
6. Carney MM, Buttell F. Reducing juvenile recidivism: evaluating the wraparound services model. *Research on Social Work Practice*. 2004;13:551–568.
7. Pullman MD, Kerbs J, Koroloff N, et al. Juvenile offenders with mental health needs: reducing recidivism using wraparound. *Crime and Delinquency*. 2006;52:375–397.
8. Kamradt B. Wraparound Milwaukee: aiding youth with mental health needs. *Juvenile Justice*. 2000;7:14–23.
9. Anderson JA, Wright ER, Kooreman HE, et al. The Dawn project: a model for responding to the needs of young people with emotional and behavioral disabilities and their families. *Community Mental Health Journal*. 2003;39:63–74.
10. Bruns EJ, Rast J, Walker JW, et al. Applying evaluation data and the wraparound process to achieve systems change for children and families. *American Journal of Community Psychology*. 2006;38:201–212.
11. Faw L. The state wraparound survey. In: Burns BJ, Goldman SK, eds. *Systems of Care: Promising Practices in Children's Mental Health, 1998 series. Vol. IV: Promising Practices in Wraparound for Children with Severe Emotional Disorders and Their Families*. Washington, DC: Center for Effective Collaboration and Practice, American Institutes for Research; 1999:27–32.
12. Weisz JR, Sandler IN, Durlak JA, et al. A proposal to unite two different worlds of children's mental health. *American Psychologist*. 2006;61:644–645.
13. US Public Health Service. Mental Health. *A Report of the Surgeon General. US Department of Health and Human Services Administration, Center for Mental Health Services*. Rockville, MD: National Institute of Mental Health; 1999.
14. US Public Health Service. *Youth Violence: A report of the Surgeon General. US Department of Health and Human Services Administration, Center for Mental Health Services*. Rockville, MD: National Institute of Mental Health; 2001.
15. US Department of Health and Human Services. Cooperative Agreements for the Comprehensive Community Mental Health Services for Children and Their Families Program (SM-05–010); 2005. Available at http://www.systemsofcare.samhsa.gov. Accessed September 2, 2005.
16. Friedman R, Drews D. Evidence based practices, systems of care, and individualized care. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, Department of Child and Family Studies. The Research and Training Center for children's Mental Health; 2005.
17. Tolan P, Dodge K. Children's mental health as a primary care and concern: a system for comprehensive support and services. *American Psychologist*. 2005;60:601–614.
18. Bruns EJ, Suter JC, Leverentz-Brady KM. Relations between program and system variables and fidelity to the wraparound process for children and families. *Psychiatric Services*. 2006;57:1586–1593.
19. Bruns EJ, Burchard JD, Suter JC, et al. Adherence to wraparound principles and association with outcomes. *Journal of Child and Family Studies*. 2005;14:521–534.
20. Ferguson C. California's title IV-E child welfare waiver demonstration project evaluation: an analysis of wraparound in Alameda County. In: Newman C, Liberton C, Kutash K, et al, eds. *The 17th Annual Research Conference Proceedings: A System of Care for Children's Mental Health*. Tampa, FL: University of South Florida, Florida Mental Health Institute Research and training Center for Children's Mental Health; 2004:149–153.
21. Hagen M, Noble K, Schick C, et al. The relationship between fidelity to wraparound and positive behavior outcomes. In: Newman C, Liberton CJ, Kutash K, et al, eds. *The 16th Annual Research Conference Proceedings, A System of Care for Children's Mental Health: Expanding the Research Base*. Tampa, FL: University of South Florida, The Louis de la Parte Florida Mental Health Institute, Research and Training Center for Children's Mental Health; 2003:293–296.

22. Suter JC, Burchard JD, Bruns EJ, et al. *User's Manual to the Wraparound Fidelity Index 3.0*. Burlington, VT: University of Vermont Department of Psychology; 2005.
23. Sather A, Bruns EJ, Stambaugh LF, et al. *The state wraparound survey. Presented at the Building on Family Strengths Conference*. Portland, OR: Research and Services in Support of Children and their Families; 2007.
24. Peterson CR, Gruner L, Earnest L, et al. Comparing functional outcomes of wraparound and traditional mental health and child welfare services. In: Newman C, Liberton CJ, Kutash K, et al, eds. *The 16th Annual Research Conference Proceedings, A System of Care for Children's Mental Health: Expanding the Research Base*. Tampa: FL: University of South Florida, The Louis de la Parte Florida Mental Health Institute, Research and Training Center for Children's Mental Health; 2004:307–311.
25. Bruns EJ, Suter JD, Leverentz-Brady K. A national portrait of wraparound implementation. In: Newman C, Liberton C, Kutash K, et al, eds. *The 16th Annual Research Conference Proceedings: A System of Care for Children's Mental Health*. Tampa, FL: University of South Florida, Florida Mental Health Institute Research and training Center for Children's Mental Health; 2004:281–286.
26. Hodges K. *Child and Adolescent Functional Assessment Scale*. Ypsilanti, MI: Eastern Michigan University, Department of Psychology; 1997.
27. Rast J, Peterson CR, Earnest L, et al. Service process as a determinant of treatment effect—the importance of fidelity. In: Newman C, Liberton CJ, Kutash K, et al, eds. *The 16th Annual Research Conference Proceedings, a System of Care for Children's Mental Health: Expanding the Research Base*. Tampa, FL: University of South Florida, The Louis de la Parte Florida Mental Health Institute, Research and Training Center for Children's Mental Health; 2004:311–315.
28. Rast J, VanDenBerg J. Certification of facilitators as a method for increasing wraparound fidelity. In: Newman C, Liberton CJ, Kutash K, et al, eds. *The 16th Annual Research Conference Proceedings, a System of Care for Children's Mental Health: Expanding the Research Base*. Tampa, FL: University of South Florida, The Louis de la Parte Florida Mental Health Institute, Research and Training Center for Children's Mental Health; 2004:315–320.
29. Walker JS, Bruns EJ. The wraparound process: Individualized care planning and management for children and families. In: Rosenberg S, Rosenberg J, eds. *Community Mental Health Reader: Current Perspectives*. New York: Routledge; 2006:44–54.
30. Walker JS, Bruns EJ. Surveying experts to define the wraparound process for children and families: An example of practice-based evidence. *Psychiatric Services*. 2006;57:1579–1585.
31. Stambaugh LF, Mustillo SA, Burns BJ, et al. Outcomes from wraparound and multisystemic therapy in a Center for Mental Health Services system-of-care demonstration site. *Journal of Emotional and Behavioral Disorders*. 2007;17:143–155.
32. Henggeler SW, Schoenwald S, Rowland MD, et al. *Serious Emotional Disturbance In Children and Adolescents: Multisystemic Therapy*. New York: Guilford; 2002.
33. Achenbach TM. *Manual for the Child Behavior Checklist/4–18 & 1991 Profile*. Burlington: University of Vermont Department of Psychiatry; 1991.
34. Stroul B, Friedman R. The system of care concept and philosophy. In: Stroul B, ed. *Children's Mental Health: Creating Systems of Care in a Changing Society*. Baltimore, MD: Paul H. Brookes; 1996.
35. Fixen DL, Naoom SF, Blasé KA, et al. *Implementation research: A synthesis of the literature. FMHI pub. no. #231*. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network; 2005.
36. Rosenheck RA. Organizational process: A missing link between research and practice. *Psychiatric Services*. 2001;52:1607–1612.