UCSF
UC San Francisco Previously Published Works

Title

Development and evaluation of a treatment fidelity instrument for family-based treatment of adolescent anorexia nervosa

Permalink

https://escholarship.org/uc/item/6b4894b3

Journal

International Journal of Eating Disorders, 48(1)

ISSN

0276-3478

Authors

Forsberg, Sarah
Fitzpatrick, Kathleen Kara
Darcy, Alison
et al.

Publication Date

2015

DOI

10.1002/eat.22337

Peer reviewed

Development and evaluation of a treatment fidelity instrument for family-based treatment of adolescent anorexia nervosa

Sarah Forsberg[a], **Kathleen Kara Fitzpatrick**[a], **Alison Darcy**[a], **Vandana Aspen**[a], **Erin C. Accurso**[b], **Susan W. Bryson**[a], **Stewart Agras**[a], **Katherine D. Arnow**[a], **Daniel Le Grange**[b], and **James Lock**[a]

[a]Department of Psychiatry and Behavioral Medicine, Stanford University School of Medicine, 401 Quarry Road, Stanford, CA 94305.

[b]Department of Psychiatry and Behavioral Neuroscience, The University of Chicago, 5841 S. Maryland Ave., Chicago, IL 60637

Abstract

Objective—The current study provides data on the psychometric properties of a newly developed measure of treatment fidelity in Family-Based Treatment (FBT) for adolescent anorexia nervosa (AN). The Family Therapy Fidelity and Adherence Check (FBT-FACT) was created to evaluate therapist adherence and competency on the core interventions in FBT.

Methods—Participants were 45 adolescents and their families sampled from three randomized clinical trials evaluating treatment for AN. Trained fidelity raters evaluated 19 therapists across 90 early session recordings using the FBT-FACT. They also rated an additional 15 session 1 recordings of an alternate form of family therapy – Systemic Family Therapy (SFT) for the purpose of evaluating discriminant validity of the FBT-FACT. The process of development and the psychometric properties of the FBT-FACT are presented.

Results—Overall fidelity ratings for each session demonstrated moderate to strong inter-rater agreement. Internal consistency of the measure was strong for sessions 1 and 2 and poor for session 3. Principal components analysis suggests sessions 1 and 2 are distinct interventions.

Conclusion—FBT-FACT demonstrates good reliability and validity as a measure of treatment fidelity in the early phase of FBT.

Keywords

anorexia nervosa; family based treatment; treatment fidelity measurement; psychometric assessment

---

Treatment fidelity is defined as the extent to which a therapeutic intervention is delivered as intended, and its assessment is critical in order to evaluate the validity of inferences drawn regarding treatment effects. Despite being a key variable in outcome research, evaluation of treatment fidelity remains limited.(1-5.) In a 2007 review of 147 child and adult randomized

Corresponding Author: Sarah Forsberg: sforsberg@stanford.edu Stanford University School of Medicine, Department of Psychiatry and Behavioral Sciences, 401 Quarry Road, Stanford, CA 94305, USA. Ph: +1 650 725 6027 Fax: +1 650 721 3322..

controlled trials (RCTs) only 3.5% were described as adequately measuring fidelity (e.g. adequately establishing, assessing, evaluating and reporting treatment integrity).(6.) Furthermore, only a handful of studies examined fidelity in psychotherapy studies involving children and adolescents or in family therapy.(7-13.) In eating disorders treatment studies we found only four RCTs that provided data on treatment fidelity.(14-17.)

The most common definition of fidelity focuses on *adherence* (defined as the utilization of specific procedures).(1; 3.) More recently, *competence* (defined as the level of skill or quality in delivering these procedures) has been added as a vital component of treatment fidelity. Additionally, fidelity relates to a specific treatment's differentiation from another treatment (i.e., proscribed practices).(1; 18.) There are several key variables that affect fidelity, including type of treatment (degree of difficulty, length, skills needed),(1.) patient characteristics (degree of pathology, motivation, readiness),(19.) therapist characteristics (training, experience, acceptability of the treatment, therapist/patient match)(1; 5.), and contextual issues that impact delivery (e.g., cost, resources needed, space needed, administrative support).(1; 4; 8.)

While several studies provide support for the usefulness of family-based treatment (FBT) for adolescents with short duration anorexia nervosa (AN), only one report on therapist fidelity to the manualized form of treatment has been published.(14.) This study used a fidelity measure developed for FBT that assessed adherence only and for which no psychometric data are provided. The authors reported that therapist fidelity varied across treatment phases, with relatively good adherence in early treatment, which diminished in middle and late treatment. The report did not evaluate specific components of FBT purported to lead to clinical success (e.g., promoting parental alignment, externalization, use of a family meal, and psycho-education about the life threatening nature of AN), nor did it directly evaluate the skill or competency with which these interventions were carried out.

Determining how best to assess fidelity in psychosocial treatments requires attention to several methodological weaknesses highlighted in the literature. First, procedures for evaluating fidelity using appropriate tools are highly variable. Further, the question of how best to assess competence (i.e. when and how often it must be assessed), will impact the nature of how fidelity and outcome are likely to be related over a course of treatment.(5; 20-22.) At the same time, ceiling effects of highly trained and supervised clinicians likely reduce variance on fidelity and therefore the likelihood of finding effects in RCTs.(23.) In addition, differences in the timing and measurement of outcome may also affect findings, with some studies suggesting early effects(24.), and others later ones.(8-10.) The development of a measurement tool that accurately and consistently captures treatment fidelity is needed to examine the impact of such variables on the fidelity-outcome relationship. Thus, the primary aim of this study is to describe the psychometric properties of an instrument designed to assess fidelity to early sessions of FBT.

Methods

Therapist and Patient Participants

Data for this study were drawn from three RCTs (two of which were multi-site studies) examining FBT for adolescents with AN. (25-27.) Study 1 was conducted at Stanford University and included 86 adolescents ages 12-18 randomized to different doses of FBT (10 sessions/6 months) or (20 sessions/12 months) (25.). Study 2 was conducted at the University of Chicago and Stanford University and included 121 adolescents ages 12-18 randomized to either FBT, or adolescent focused therapy, an individual treatment for AN (26.). Study 3 included 164 adolescents from a seven-site international RCT comparing FBT to Systemic Family Therapy (SFT) (27.). For each study, research was reviewed and approved by an institutional review board, with adult participants providing consent and child participants their assent. The final sample (N=30) was randomly selected from a total of 371 patient participants based on availability of three consecutive selected audio or video session recordings (sessions 1, 2 and 3). The SFT sample of session 1 recordings (N=15) was randomly selected from Study 3. Participant demographics and baseline (BL) characteristics are described in detail elsewhere. (25-27.) Briefly, they were adolescents (male and female) ages 12-18 with a DSM-IV-TR diagnosis of AN, without requiring the 3 month loss of menses criteria. There were 10 participants from Study 1, 4 from Study 2, and 16 from Study 3.

Therapists whose therapy sessions were evaluated in this study included doctoral level psychologists, psychiatrists, masters' level family therapists, and doctoral students in psychology. The final sample of therapists whose session recordings were rated included 14 FBT therapists and 5 SFT therapists from 3 studies. (25-27.) Years of experience treating adolescents with AN ranged from 2-20 years, and experience with FBT ranged from 1-12 years. Therapists were trained using standardized training procedures, which included an in-person two-day training by the authors of the manual who subsequently directly reviewed audio and video-taped therapy sessions prior to therapist approval to treat randomized participants. The authors of the manual (JL and DLG) provided weekly group supervision thereafter. In Study 1, JL provided weekly in-person supervision and in Study 2, JL and DLG both provided in-person supervision at their respective sites. In Study 3, trained FBT supervisors provided on-site training, and training was also conducted across site, with continuous review of videotaped sessions by an expert in FBT (KF).

Fidelity Measures

The current fidelity instrument was created through expansion and enhancement of a previously developed measure (the Family Therapy Fidelity Check) for use in assessing therapist adherence and competence in an RCT. The Family Therapy Fidelity and Adherence Check (FBT-FACT) was established for the purposes of rating early therapy sessions (1, 2 and 3-8), and items varied depending on interventions associated with particular sessions (Table 1). Good treatment outcome in FBT has been linked to early change (e.g., weight gain by session 4), providing a rationale for emphasizing fidelity to the model early in treatment. (28; 29) The FBT-FACT was designed to both assess the presence/absence of a treatment goal and, if implemented, the quality of the intervention. Per the

treatment manual, items chosen for assessment map directly on to treatment goals and interventions for each session. Only therapist behaviors were rated since the goal of the measure was to evaluate the presence of specific behaviors and the skill of implementation. Interventions were coded as "not applicable" when they were inappropriate (e.g., "aligning the patient with siblings" when no siblings were present in the session). Additionally, codes were rated for their fidelity and intent, rather than their outcome. For example, therapists received high ratings for skillful implementation of a strategy (e.g., introducing externalization), regardless of the family's ability to grasp the concept. Certainly, higher ratings were given to therapists who appropriately expanded on externalization in response to a family's difficulty in understanding.

Competence was established by scoring the effectiveness of the therapist at accomplishing specific intervention goals. Competence was rated on a 7-point Likert scale with effectiveness rated from a score from 1= "not at all" to 7= "very much." Finally, an item assessing overall fidelity for each session was rated on a Likert scale where 1= "minimal fidelity" to 7 = "excellent fidelity / excellent fluid session / excellent use of skills and goals met." A fidelity manual was written in conjunction with a coding framework to enhance understanding of meaning underlying different therapeutic interventions in FBT and to anchor the coding framework, distinguishing a "1" from a "4" and a "7", and so on. These anchors were developed to reflect the types of behaviors (or lack thereof) that would be coded at each value of the seven-point scale but were not meant to be all-inclusive.

To elucidate the use of the seven-point scale, an illustration for benchmark anchors associated with the fidelity item assessing *Weight Feedback* is provided. At the beginning of every session, therapists implementing FBT are expected to take and share the patient's weight with the family. Goals and examples associated with this particular task were created for each benchmark score to ensure shared understanding and objectivity in ratings. On this item, a high score of 7 was described as reflecting the following therapist behaviors "Provides feedback at the start of the session, including both verbal and graphical representation of the change; explaining to parents the purpose of the weigh-in and the way this information will be used; discourages weigh-ins outside of session; provides education about the purpose of the weigh-in (to provide exposure to weight for patient; to allow for assessment of progress toward goals, to activate AN in the session); and, provides specifics regarding expected weight progress (e.g., 2 pounds a week)". A benchmark description for a score of 1 on this item was written as follows – "Therapist response is in direct contrast to the weight (a concerned, heavily problem oriented response to an appropriately increased weight, which should otherwise be met with congratulations OR a cheerful, chatty response to weight loss); OR weight is taken and not shared". This coding framework reflects an effort to operationalize therapist competence, given that the previously used measure only rated adherence (i.e., presence or absence of a particular therapist behavior).

The final measure also includes separate items for sessions 1, 2 and 3-8 (phase I) of FBT, where some therapist tasks are meant to occur across sessions (e.g. externalization of illness), and others are unique to a specific session (e.g. taking a history of AN in session 1). For the purposes of this study, only sessions 1, 2 and 3 were rated. Table 1 provides a reference to fidelity item descriptions for each session.

Training of Judges/Rating Procedures

Three doctoral level psychologists with 3.5 years to 8 years of experience with FBT for AN completed ratings of therapy sessions. The lead coder (KF) was a supervisor and rater on the large multi-site study and had substantial experience rating recordings for fidelity using the initial measure and was responsible for training all coders on the FBT-FACT. During the process of rating the recordings using the initial measure, note taking identified the challenges and range of behaviors that were missed using the earlier adherence measure.

The use of a rating scale for skill of intervention was developed and 25 recordings were re-rated using this measure. The goal was to identify whether the scaling (1-7) was appropriate to capture the range of skills in implementing these interventions and to identify challenges in coding. It was during this process that the "Not applicable" code was introduced. An initial coding framework provided in-depth descriptions for benchmark codes. The lead coder also addressed common challenges, such as understanding the relationship between family members (identification of siblings, parents versus step-parents). In addition, one item added to the measure included "overall competence/fidelity" when implementing these tasks. When done well, the first session of FBT builds in intensity and momentum, culminating with the introduction of the task of re-nourishment and charging the parents with the family meal. The importance of fluidity in successive introduction of interventions was also considered a proxy measure for the therapist's comfort with implementing these skills and adherence to the model.

Following the development of the initial coding manual, the three raters evaluated the same six recordings (20% of the total sample [N=30]). These were watched as a group when possible, but coded independently, then discussed to identify challenges in reaching consensus on particular items. The lead coder then compiled these responses and developed a coding framework for the FBT-FACT that provided anchors for the likert-type ratings. This coding framework was then collaboratively refined.

Statistical analyses

Reliability—Inter-rater reliability was examined by calculating an intra class correlation (ICC) using a twoway random effect model as well as Spearman-Brown correction representing the mean reliability across the two raters according to recommendations of Shrout and Fleiss.(30.) ICC is therefore a measure of how two randomly selected raters perform and was calculated for each fidelity code independently. A cutpoint ICC of >.40 was used to determine whether an item would be retained on the measure. Items .40 were not retained and excluded in subsequent analyses.

Internal consistency data—Adjusted item to scale (session average) correlations provide an estimate of the convergence between the item being evaluated and the rest of the items on the measure. In calculating the adjusted item to scale correlations, the item to be evaluated was excluded from the total to avoid inflating the correlation. An item was considered to possess adequate convergence if its adjusted item-to-scale correlation was .30.(31.) We also calculated Cronbach's alpha coefficients as a measure of internal consistency using a threshold of .70 as a standard for reasonable internal reliability.(31.)

Principal Components Analysis—A maximum likelihood principal components analysis (PCA) with varimax rotation was conducted to explore latent constructs. We used Horn's parallel analysis (32.) and Velicer's minimum average partial analysis (33.) to determine the number of components to extract. These methods were chosen over standard approaches like Kaiser's eigenvalue > 1 rule, and Cattell's scree plot test given these methods are thought to inconsistently identify or overestimate the number of components to extract. We utilized syntax detailed in O'Conner to determine the number of components. (34.) The PA involves comparisons between the actual dataset and randomly generated datasets with the same number of observations and variables as in the original dataset. Eigenvalues are then extracted from the random datasets, and are compared to the eigenvalues generated in the initial PCA. Eigenvalues that equate to the 95th percentile are compared to those from the observed data and values greater than the random data are retained.

Prior to running PCA and internal reliability analyses, inter-item variability was assessed to determine whether data met assumption of normality. Those items that were skewed were removed from subsequent analyses (*Therapist Agnosticism* for all sessions, and *Sibling Support* for sessions 2 and 3), leaving seven items for session 1, and nine items for session 2. Where we planned to include all session items in the analysis, only sessions 1 and 2 were analyzed together as session 3 items demonstrated poor internal consistency.

Discriminant validity—To determine whether the FBT-FACT is capable of distinguishing FBT from other forms of family therapy, we utilized the measure to rate fidelity of Systemic Family Therapy (SFT). Two raters were randomly selected to rate 10 recordings each (total of 15 session recordings). Session 1 recordings were compared in this analysis given that session 2 recordings would be expected to be vastly different between treatments (i.e., session 2 in FBT is the family meal where in SFT this does not occur). Contrast analyses (independent-samples t-test for available competency ratings) and descriptive analyses (adherence frequencies) were completed for 15 session 1 recordings.

Results

Descriptive results

Many of the items had perfect or near-perfect adherence ratings (see Table 2 for descriptive data on adherence and competence). Means and standard deviations (SDs) of competency ratings were calculated with data coded as missing for those items where therapist adherence was coded as nonexistent ("No" or "NA"). Items with low (<50%) adherence were *Therapist Agnosticism* in sessions 2 and 3, *Modification of Criticism* in session 2 and *Sibling Support* in session 2. Means and SDs can be seen in Table 2. For all subsequent analyses, a score of 0 was assigned those cases that were non-adherent, extending the Likert scale from 0 (non-adherent) to 7 (perfect fidelity).

Interrater Reliability

For ratings of therapist competence, interrater reliability (intra class correlation [ICC]) ranged from -.12 to .94 (see Table 2). There were a number of items on the measure with

poor ICCs (*Charge with Re-feeding* in session 1; *Modification of Criticism* in sessions 1, 2 and 3; *Therapist Agnosticism* in session 2, and *Focus on Eating* and *Weight Feedback* in session 3). As a result, these items were excluded in subsequent analyses. Interrater agreement for an item assessing *Overall Fidelity* was moderate to strong (ICC=[.61 -.77]).

Internal consistency

Calculation of Cronbach's alpha for each session revealed a high level of internal consistency of items for sessions 1 (no. items=7; alpha=.867) and 2 (no. items=9; alpha=.827). After excluding *Therapist Agnosticism* (sessions 1 and 2) and *Sibling Support* (session 2) due to non-normal distribution, session 3 consisted of 4 items and internal consistency was poor (alpha=.433). The only item with low item-total correlations was *Sibling Support* in session 2. Removal of this item would have resulted in higher degree of internal consistency within session 2.

Principle Components Analysis

The following items were included in the PCA for sessions 1 and 2 after *Therapist Agnostic* sessions 1 and 2, and *Sibling Support* session 2 were removed for non-normal distribution: Session 1: *Greet the Family*, *Externalization*, *Orchestrate Intense Scene*, *Reduce Parental Guilt*, *History of AN*, *Family History* and Session 2: *Weight Feedback*, *History of Eating Pattern*, *Externalization*, *Nutritional Needs*, *Align Parents in Renourishment*, *One More Bite*, *Best Re-feed*, and *Focus on AN*. The final rotated component matrix showed that as intended, the measure can be represented by two components representing session one and session two. Only one session 2 item *Weight Feedback* loaded onto both components, and the item was retained for conceptual reasons as a component of session 2. This 14-item solution explained 63% of the variance (Table 3).

Item-Overall Fidelity correlations

Items with high factor loadings were subjected to correlational analysis to determine if these items could be subsumed under an item assessing *Overall Fidelity*. There was a strong positive relationship between *Overall Fidelity* and Session 1 and 2 items (see Table 2).

Discriminant validity

The FBT-FACT successfully discriminated FBT from SFT on the majority of items, specifically those that are considered unique to FBT. Raters generally agreed that therapists in SFT did not deliver the following interventions specified on the measure: *Orchestrate Intense Scene*, *Reduce Parental Guilt*, *Modify Criticism* and *Charge with Re-feeding*, *Externalization* and *Therapist Agnosticism*. Only on the item *History of AN* did raters demonstrate low levels of agreement (40%) on whether the intervention had occurred. As expected, the FBT-FACT did not distinguish SFT from FBT on *Greet the Family* or *Family History* since these two components are common to both treatments. Results of independent-samples t-tests comparing competency ratings on items revealed significantly lower competency ratings in SFT on *Greet the Family* (SFT: M=2.97, SD=0.90; FBT: M=4.47, SD=1.27; t (43)=4.10, *p* <.001) and significantly higher competency ratings on *Family History* (SFT: M=5.03, SD=0.67; FBT: M=4.10, SD=1.26); t (42.76)=-3.26, *p*<.01).

Discussion

This is the first study to evaluate the psychometric properties of a fidelity measure for FBT. Therapists were highly adherent in their delivery of FBT core therapeutic interventions in the first three sessions. Scores related to the delivery of individual therapeutic interventions generally represented competence in the adequate range (i.e., textbook implementation without expansion, generalization, or utility for the family). Psychometric properties of the FBT-FACT suggest that this tool may be useful to assess both adherence and competence in FBT. Interrater reliability varied across items. Raters showed moderate to strong agreement on their assessment of *Overall Fidelity* despite variability in single-item agreement, suggesting that therapist competence across sessions can be measured with good validity despite some disagreement on item-level ratings.

Sessions 1 and 2 demonstrated high levels of internal consistency, whereas items on session 3 had poor internal consistency. While preliminary, results of the PCA in combination with strong internal reliability, supports retaining sessions 1 and 2 as distinct measurements on the FBT-FACT. On the contrary, the measure does not appear to be a reliable representation of the interventions as currently written for session 3. Taken together, results of the PCA and internal consistency analysis suggest that specific interventions nested in sessions 1 and 2 are highly related, and also that competent delivery of one therapeutic intervention is reliant on competent delivery of others. Given interventions are intended to complement and build on one another within session, the interdependence of competency ratings in session 1 and 2 is not surprising. Further examination of the relatedness of session 3 items is likely contingent on resolution of poor IRR on items thought to be central to the session (e.g. *Focus on Eating* and *Support Re-feeding*). Together with good internal validity, strong inter-item correlations, robust factor loadings, and good IRR of *Overall Fidelity*, these findings suggest that therapist competence may be effectively and efficiently rated as one global dimension.

The final ingredient in establishing treatment integrity is the extent to which the treatment of interest can be differentiated from other treatments. In this study, the FBT-FACT was able to distinguish the majority of items on FBT from SFT (e.g. therapists did not place responsibility for refeeding on parents in SFT, a central intervention in FBT). Raters differed in their perception of whether SFT therapists took a history of AN, perhaps due to differing standards on the timing, structure, and specificity with which the history should be taken. Indeed, the history of AN taken in SFT—when present—was different from that of FBT as it was structured through a genogram utilized to take general family history. There were also specific interventions that occurred at equal frequency in SFT and FBT, suggesting that the FBT-FACT includes items that relate to family therapy more broadly (i.e., greeting the family and taking a family history), even though accomplished differently in each treatment. In FBT, the focus in greeting the family is done in a grave and sincere manner to frame subsequent orchestration of an intense scene, a stylistic method that is absent from SFT. In contrast, SFT therapists had higher ratings on taking a family history, which was expected given this is the primary focus of the first session of the treatment.

Challenges in rating fidelity

This study highlights some of the challenges inherent in treatment fidelity research. Adequate rater agreement was not achieved on one-quarter of the items rated. While IRR scores may appear low in this study, they are consistent with the range of scores found in other psychometric analyses of fidelity instruments, highlighting the greater difficulty in achieving agreement on fidelity that encompasses competency ratings,(7; 35; 36.) than adherence ratings. (37; 38.) Inconsistency in rater agreement likely reflects differences in raters' conceptualization of competence, the aspects of the intervention emphasized and challenges in operationalizing competence.(5.) Fidelity evaluators in this study identified these issues as core challenges in their efforts. For example, in session 3 where some items had poor-fair interrater reliability, raters identified increased variability in the content and organization of sessions as interventions are applied flexibly depending on how the family has responded to earlier interventions. The manual allows greater flexibility to therapists in tailoring session 3 interventions to a family's needs. This is in contrast to sessions one and two, for which the manual prescribes numerous, highly structured, specific tasks. Further, randomization was conducted by session, not participant. Therefore, it is possible that ratings may have been different had they been conducted in the broader context of earlier sessions.

Challenges in achieving strong IRR also reflect raters' difficulty in quantifying poor implementation and suggests the need for better therapist training in certain interventions. For example, *Weight* Feedback in session 3, while seemingly simple, relies on timing (i.e., sharing weight proactively at the beginning of the session), specificity (i.e., sharing more specific weight change information), rationale (i.e., explanation for taking and discussing weight change), and context (i.e., how weight is used to shape the family toward improved re-feeding), for high competency. Another intervention, *Focus on Eating,* is also multifaceted, and future fidelity codes might be divided further to increase specificity (e.g., assessing and identifying barriers to re-feeding at home, expanding on successes in re-feeding, supporting development of family use of problem solving skills, reframing non-eating related challenges in context of eating disorder). Low frequency interventions may have also reduced the variability of therapist and family behaviors across which evaluators could rate. These included *Modify Criticism* (likely due in part to the relatively low frequency of overtly critical comments observed in this sample) (39; 40.) and *Therapist Agnosticism* in session 2. In considering these potential challenges in rater agreement, future efforts related to training and fidelity measurement may need to describe these interventions with greater specificity in order to obtain improved competency. Rater feedback on challenges can be incorporated into future iterations of the FBT manual. Furthermore, enhanced training and supervision protocols may help remedy inconsistencies in fidelitous implementation of specific interventions outlined above.

Limitations

An important limitation to consider when interpreting this study's findings is its relatively small sample size. The significant resources required to rate full-length therapy sessions is a well-documented barrier to conducting fidelity research.(41.) The availability of sessions in this study was limited due to the need for three intact consecutive sessions and outcome data

for each participant. It is possible that missing data was non-random, and attributable to therapist or family variables (of note, these were primarily in-tact families). There are a number of implications related to the small sample size. First, it could contribute to variability in interrater reliability scores, which may have been improved upon with exposure to wider variability of sessions. Second, sample size limits the generalizability of the factor structure analyses, and replication on a larger sample following recommended guidelines (i.e. completing a confirmatory factor analysis) is warranted. In spite of this, our findings (i.e., strong loadings of many items on each factor, generally high communalities, and only one item cross-loading) mitigate concern about sample size, consistent with published standards.(42.) Results of the PCA should be considered preliminary and taken in the context of other analyses and theoretical background presented here. Finally, we limited our ratings to early therapy sessions, given the importance of early change in FBT promoting optimal treatment outcome. (28.) Therefore, we are unable to comment on fidelity to interventions used in the second and third phases of treatment.

Strengths and Future Directions

This study of treatment fidelity is in compliance with best-practice recommendations for the field. We assessed primary psychometric properties of the measure including discriminant validity, where the minority of studies examining treatment fidelity have done so.(43.) The strengths of this study provide a solid foundation for further validation of the FBT-FACT, such as examining its predictive validity. Knowledge of fidelity in FBT is in its infancy, and this study is only a first step in this measure's evaluation, which must also include a complementary examination of other factors proposed to influence therapist fidelity. For example, therapist, patient and treatment setting are all variables that may be predictive of differences in fidelity as they have been in other populations, and should be explored.(7; 25; 44.) Ultimately, an important aim of treatment fidelity research is to refine treatment and training to improve successful dissemination and patient outcomes. Central to this goal is strengthening the measure by further operationalizing interventions to increase specificity and interrater agreement thereby enhancing confidence in our ability to disseminate the measure outside of a controlled-trial. Modifications to the manual that incorporate findings from this study may enhance training and supervision procedures such that specific elements with consistently low fidelity are emphasized and carefully monitored. These refinements throughout the process of training, monitoring, and assessment of fidelity provide a foundation from which analysis of the relationship between fidelity in FBT and treatment outcome may be carefully examined.

References

1. Perepletchikova F, Kazdin A. Treatment integrity and therapeutic change: issues and research recommendations. Clinical Psychology Science and Practice. 2005; 12:365–383.

2. McLeod B, Weisz J. Conceptual and methodological issues in treatment integrity measurement. School Psychology Review. 2009; 38:541–546.

3. Gresham F. Evolution and the treatment integrity concept: Current status and future directions. School Psychology Review. 2009; 38:533–540.

4. McHugh R, Murray H, Barlow D. Balancing fidelity and adaptation in the dissemination of empirically supported treatments: The promise of transdiagnostic interventions. Behaviour Research and Therapy. 2009; 47:946–953. [PubMed: 19643395]

5. Barber J, Sharpless B, Klostermann S, McCarthy K. Assessing intervention competence and its relation to therapy outcome: A selected review derived from the outcome literature. Professional Psychology: Research and Practice. 2007; 38:493–500.

6. Perepletchikova F, Treat T, Kazdin A. Treatment integrity in psychotherapy research: Analysis of the studies and examination of the associated factors. Journal of Consulting and Clinical Psychology. 2007; 75:829–841. [PubMed: 18085901]

7. Hogue A, Henderson C, Dauber S, Barajas P, Fried A, Liddle H. Treatment adherence, competence, and outcome in individual and family therapy for adolescent behavior problems. Journal of Consulting and Clinical Psychology. 2008; 76:544–555. [PubMed: 18665684]

8. Henggeler S, Brondino M, Melton G, Scherer D. Multisystemic therapy with violent and chronic juvenile offenders and their families: The role of treatment fidelity in successful dissemination. Journal of Consulting and Clinical Psychology. 1997; 65:821–833. [PubMed: 9337501]

9. Henggeler S, Pickrel S, Brondino M. Multisystemic treatment of substance-abusing and dependent delinquents: Outcomes, treatment fidelity and transportability. Mental Health Services Research. 1999; 1:171–184. [PubMed: 11258740]

10. Huey S, Henggeler S, Brondino M, Pickrel S. Mechanisms of change in multisystemic therapy: Reducing delinquent behavior through therapist adherence and improved family and peer functioning. Journal of Consulting and Clinical Psychology. 2000; 68:451–467. [PubMed: 10883562]

11. Weisman A, Tompson M, Okazaki S, Gregory J, Goldstein M, Rea M, et al. Clinicians' fidelity to a manual-based family treatment as a predictor of one-year course of bipolar disorder. Family Process. 2002; 41:123–131. [PubMed: 11924080]

12. Weisz J, Southam-Gerow M, Gordis E, Connor-Smith J, Chu B, Langer D, et al. Cognitive-Behavioral Therapy versus Usual Clinical Care for Youth Depression: An Initial Test of Transportability to Community Clinics and Clinicians. 2009; 77:383–396.

13. Schoenwald S, Carter R, Chapman J, Sheidow A. Therapist adherence and organizational effects on change in youth behavior problems one year after multisystemic therapy. Administration and Policy in Mental Health and Mental Health Services Research. 2008; 35:379–394. [PubMed: 18561019]

14. Couturier J, Isserlin L, Lock J. Family-based treatment for adolescents with anorexia nervosa: A dissemination study. Eating Disorders. 2010; 18:199–209. [PubMed: 20419524]

15. Wilfley D, Welch R, Stein R, Spurrell E, Cohen L, Saelens B, et al. A Randomized Comparison of Group Cognitive-Behavioral Therapy and Group Interpersonal Psychotherapy for the Treatment of Overweight Individuals With Binge-Eating Disorder. Archives of General Psychiatry. 2002; 59:713–721. [PubMed: 12150647]

16. Loeb K, Wilson G, Labouvie E, Pratt E, Hayaki J, Walsh B, et al. Therapeutic alliance and treatment adherence in two interventions for bulimia nervosa: A study of process and outcome. Journal of Consulting and Clinical Psychology. 2005; 73:1097–1107. [PubMed: 16392983]

17. Schlup B, Munsch S, Meyer A, Margraf J, Wilhelm F. The efficacy of a short version of cognitive-behavioral treatment followed by booster sessions for binge eating disorder. Behaviour Research and Therapy. 2009; 47:628–635. [PubMed: 19446793]

18. Barber J, Folz C, Crits-Christoph P, Chittams J. Therapists' adherence and competence and treatment discrimination in the NIDA Collaborative cocaine treatment study. Journal of Clinical Psychology. 2004; 60:29–41. [PubMed: 14692007]

19. Barber J, Crits-Christoph P, Luborsky L. Effects of therapist adherence and competence on patient outcome in brief dynamic therapy. Journal of Consulting and Clinical Psychology. 1996; 64:619–622. [PubMed: 8698958]

20. Trepka C, Rees A, Shapiro D, Hardy G, Barkham M. Therapist competence and outcome of cognitive therapy for depression. Cognitive Therapy and Research. 2004; 28:143–157.

21. Stiles W, Honos-Webb L, Surko M. Responsiveness in psychotherapy. Clinical Psychology Science and Practice. 1998; 5:429–458.

22. Jacobson N, Dobson K, Traux P, Addis M, Koerner K, Gollan J, et al. A component analysis of cognitive-behavioral treatment for depression. Journal of Consulting and Clinical Psychology. 1996; 64:293–304.

23. Whisman M. Moderators and mediators of change in cognitive therapy of depression. Psychological Bulletin. 1993:114.

24. Shaw B, Olmsted M, Dobson K, Sotsky S, Yamaguchi J, Elkin I, et al. Therapist competence ratings in relation to clinical outcome in cognitive therapy of depression. Journal of Consulting and Clinical Psychology. 1999; 67:837–846. [PubMed: 10596506]

25. Lock J, Agras WS, Bryson S, Kraemer HC. A Comparison of Short- and Long-Term Family Therapy for Adolescent Anorexia Nervosa. Journal of the American Academy of Child & Adolescent Psychiatry. 2005; 44:632–639. [PubMed: 15968231]

26. Lock J, Le Grange D, Agras W, Moye A, Bryson S, Jo B. Randomized clinical trial comparing family-based treatment with adolescent-focused individual therapy for adolescents with anorexia nervosa. Archives of General Psychiatry. 2010; 67:1025–1032. [PubMed: 20921118]

27. Lock J, Brandt H, Woodside B, Agras S, Halmi W, Johnson C, et al. Challenges in conducting a multi-site randomized clinical trial comparing treatments for adolescent anorexia nervosa. International Journal of Eating Disorders. 2012; 45:202–213. [PubMed: 21495052]

28. Doyle P, Le Grange D, Loeb K, Doyle A, Crosby R. Early response to family-based treatment for adolescent anorexia nervosa. International Journal of Eating Disorders. 2010; 43:659–662. [PubMed: 19816862]

29. Le Grange D, Accurso E, Lock J, Agras S, Bryson S. Early weight gain predicts outcome in two treatments for adolescent anorexia nervosa. International Journal of Eating Disorders. 2014; 47:124–129. [PubMed: 24190844]

30. Shrout P, Fleiss J. Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin. 1979; 86:420–428. [PubMed: 18839484]

31. Nunnally J, Bernstein I. Psychometric Theory. New York: McGraw-Hill. 1994

32. Horn J. A rationale and test for the number of factors in factor analysis. Psychometrika. 1965; 30:179–185. [PubMed: 14306381]

33. Velicer W. Determining the number of components from the matrix of partial correlations. Psychometrika. 1976; 41:321–327.

34. O'Connor B. SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. Behavior Research Methods, Instrument Computers. 2000; 32:396–402.

35. Barber J, Liese B, Abrams M. Development of the Cognitive Therapy Adherence and Competence Scale. 2003; 13:205–221.

36. Hogue A, Dauber S, Chinchilla P, Fried A, Henderson C, Inclan J, et al. Assessing fidelity in individual and family therapy for adolescent substance abuse. Journal of Substance Abuse Treatment. 2008; 35:137–147. [PubMed: 17997268]

37. Alvarez-Jimenez M, Wade D, Cotton S, Gee D, Pearce T, Crist K, et al. Enhancing treatment fidelity in psychotherapy research: Novel approach to measure the components of cognitive behavioral therapy for relapse prevention in first-episode psychosis. Australian and New Zealand Journal of Psychiatry. 2008; 42:1013–1020. [PubMed: 19016089]

38. Weck F, Bohn C, Ginzburg D, Stangier U. Assessment of adherence and competence in cognitive therapy: Comparing session segments with entire sessions. Psychotherapy Research. 2011; 21:658–669. [PubMed: 21793688]

39. Hoste R, Labuschagne Z, Lock J, Le Grange D. Cultural variability in expressed emotion among families of adolescents with anorexia nervosa. International Journal of Eating Disorders. 2012; 45:142–145. [PubMed: 22170027]

40. Le Grange D, Hoste R, Lock J, Bryson S. Parental expressed emotion of adolescents with anorexia nervosa: Outcome in family-based treatment. International Journal of Eating Disorders. 2011; 44:731–734. [PubMed: 22072411]

41. Perepletchikova F, Chereji E, Hilt L, Kazdin A. Barriers to implementing treatment integrity procedures: Survey of treatment outcome researchers. Journal of Consulting and Clinical Psychology. 2009; 77:212–218. [PubMed: 19309181]

42. Costello A, Osborn J. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. Practical Assessment, Research & Evaluation. 2005; 10:1–9.

43. Schoenwald S, Garland A. A review of treatment adherence measurement methods. Psychological Assessment. 2013; 25:146–156. [PubMed: 22888981]

44. Le Grange D, Lock J, Agras W, Moye A, Bryson S, Jo B, et al. Moderators and mediators of remission in family-based treatment and adolescent focused therapy for anorexia nervosa. Behavior Research and Therapy. 2012; 50:85–92.

Table 1

FBT Fidelity and Adherence Check (FBT-FACT) items

Item	Abbreviation	Session 1	Session 2	Session 3
Did the therapist greet the family in a sincere but grave manner?	Greet the Family	X		
Did the therapist take a history that engages each family member in the process?	Family History	X		
Did the therapist gather a history focused on AN, rather than collect a general history?	History of AN	X		
Did the therapist separate the illness from the patient (e.g., through the use of metaphors, Venn diagrams)?	Externalization	X	X	X
Did the therapist orchestrate an intense scene around the seriousness of the illness and difficulty in recovery (raising parental motivation)?	Orchestrate Intense Scene	X		
Did the therapist assist the family in reducing guilt and/or blame?	Reduce Parent Guilt	X		
Did the therapist remain agnostic to the cause of AN?	Therapist Agnosticism	X	X	X
[a]Did the therapist modify parental and sibling criticisms (if present)?	Modify Criticism	X	X	X
[a]Did the therapist charge the parents with the task of re-feeding?	Charge with Re-feeding	X		
Did the therapist provide feedback to the patient and family regarding weight?	Weight Feedback		X	X
Did the therapist take a history of the family patterns around food preparation, food serving, and family discussions about eating especially as it relates to the patient?	History of Eating Pattern		X	
Did the therapist assist the family in understanding nutritional needs of the patient (if necessary)?	Nutritional Needs		X	
[a]Did the therapist align the parents in an effort to work together in regard to renourishment of the patient?	Align Parents in Renourishment		X	
Did the therapist help the parents convince their child to eat at least one mouthful more than s/he was prepared to?	One More Bite		X	
Did the therapist help set the parents on their way to work out among themselves how best they can go about in re-feeding their child?	Best Re-feed		X	
Did the therapist work to align the patient with her siblings for support?	Sibling Support		X	X
Did the therapist keep the focus on AN and eating disorder behavior?	Focus on AN		X	
[a]Did the therapist direct, redirect, and focus therapeutic discussion on food and eating behaviors and their management until food, eating, and weight behaviors and concerns are relieved?	Focus on Eating			X
Did the therapist discuss, support, and help parental dyad's efforts at re-feeding?	Support Re-feeding			X
Overall, how would you rate the fidelity of this recording?	Overall Fidelity	X	X	X

Note.

[a]Item removed due to low ICC; X=intervention should occur in the session per the FBT manual.

Table 2

Interrater reliability and fidelity descriptives

Session items	Adherence	Competence	Inter-rater reliability	Correlation[b]
Session 1	**%**	**M (SD)**	**ICC**	**r**
Greet the Family	100	4.47 (1.27)	.72	.80[*]
Family History	100	4.10 (1.26)	.49	.78[*]
History of AN	100	4.40 (1.37)	.45	.80[*]
Externalization	100	3.98 (1.33)	.80	.83[*]
Orchestrate Intense Scene	100	4.03 (1.56)	.70	.67[*]
Reduce Parent Guilt	87	3.10 (1.54)	.46	
Therapist Agnosticism	63	3.13 (1.52)	.72	
Modify Criticism	80	1.98 (1.18)	.33	
Charge with Re-feeding	87	4.00 (1.10)	.08[a]	
Overall Fidelity	100	4.18 (1.19)	.61	

Session items		Adherence	Competence	ICC	r
Session 2		**%**	**M (SD)**	**ICC**	**r**
Externalization		97	3.95 (1.33)	.76	.60[*]
Therapist Agnosticism		10	2.50 (1.80)	.23[a]	
Modify Criticism		33	2.55 (1.34)	.36[a]	
Weight Feedback		73	3.30 (1.59)	.83	
History of Eating		100	4.42 (.97)	.43	.54[*]
Nutritional Needs		97	3.66 (1.39)	.68	.74[*]
Align Parents in Renourishment		93	2.91 (1.63)	.77	.56[*]
One More Bite		90	3.43 (1.45)	.74	.70[*]
Best Re-feed		100	3.95 (1.23)	.57	.70[*]
Sibling Support		43	3.27 (1.70)	.94	
Focus on AN		100	5.20 (1.16)	.73	.67[*]
Overall Fidelity		100	4.13 (1.18)	.72	

Session items	Adherence	Competence	ICC	
Session 3				
Weight Feedback	100	4.27 (.73)	−.12[a]	
Sibling Support	57	3.50 (1.26)	.90	
Focus on Eating	100	4.45 (1.02)	.38[a]	
Support Re-feeding	100	3.85 (1.02)	.60	
Therapist Agnosticism	7	3.00 (2.12)	.57	
Externalization	100	4.12 (1.23)	.49	
Modify Criticism	63	3.34 (1.75)	.40[a]	

Session 3

Overall Fidelity	100	4.18 (1.11)	.77

Note.

[a] item excluded due to low ICC. ICC cutpoints: 0-.2=poor, .3-.4=fair, .5-.6=moderate, .7-.8-strong, >.8=almost perfect

[b] Correlation between item and Overall Fidelity item. Items included in correlational analysis are those with high factor loadings in PCA.

[*] $p < .001$

[*] $p < .01$.

Table 3

Summary of PCA results for fidelity items

	Item	Factor 1	Factor 2
Session 1	Greet the Family	**.902**	.091
	Family History	**.878**	.119
	History of AN	**.879**	.088
	Externalization	**.846**	−.027
	Orchestrate Intense Scene	**.702**	−.272
	Reducing Guilt	**.659**	.056
Session 2	Weight Feedback	−.504	**.554**
	History Eating Pattern	.252	**.698**
	Nutritional Needs	−.059	**.872**
	Renourishment	.084	**.684**
	One More Mouthful	−.123	**.755**
	Best Re-feed	−.082	**.795**
	Externalization	.205	**.651**
	Focus on AN	−.101	**.802**
Eigenvalue		4.52	4.29
Percent variance		31.44	31.43

Table 4

Comparison of FBT to SFT and Discriminant Validity of the FBT-FACT

	FBT		SFT	
Item	**Rater agreement (%)**	**Adherence (%)**	**Rater agreement (%)**	**Adherence (%)**
Greet Family	100%	100% (N=30)	80%	100% (N=12)
Family History	93%	100% (N=28)	100%	100% (N=15)
History of AN	100%	100% (N=30)	40%	33% (N=6)
Externalization	97%	100% (N=29)	93%	0% (N=14)
Orchestrate Intense Scene	87%	100% (N=26)	100%	0% (N=15)
Reduce Parent Guilt	73%	82% (N=22)	100%	7% (N=15)
Therapist Agnostic	77%	52% (N=23)	73%	0% (N=11)
Modify Criticism	50%	60% (N=15)	100%	0% (N=15)
Charge with Re-feeding	93%	93% (N=28)	100%	0% (N=15)

Note: See Table 1 for description of each item.