

NIH Public Access

Author Manuscript

J Consult Clin Psychol. Author manuscript; available in PMC 2014 August 01.

Published in final edited form as:

J Consult Clin Psychol. 2013 August ; 81(4): 674–680. doi:10.1037/a0033021.

Comparison of Youth, Caregiver, Therapist, Trained, and Treatment Expert Raters of Therapist Adherence to a Substance Abuse Treatment Protocol

Jason E. Chapman,

Department of Psychiatry and Behavioral Sciences, Medical University of South Carolina

Michael R. McCart,

Department of Psychiatry and Behavioral Sciences, Medical University of South Carolina

Elizabeth J. Letourneau, and Department of Mental Health, Bloomberg School of Public Health, Johns Hopkins University

Ashli J. Sheidow

Department of Psychiatry and Behavioral Sciences, Medical University of South Carolina

Abstract

Objective—This study evaluated the accuracy of youth, caregiver, therapist, and trained raters relative to treatment experts on ratings of therapist adherence to a substance abuse treatment protocol for adolescents.

Method—Adherence ratings were provided by youth and caregivers in an ongoing trial evaluating a Contingency Management (CM) intervention for youth in juvenile drug court. These ratings were compared to those provided by therapists and trained raters, and each rater type was compared to ratings provided by CM treatment experts. Data were analyzed using IRT-based Many-Facet Rasch Models.

Results—Relative to treatment experts, youth and caregivers were significantly more likely to endorse the occurrence of CM components. In contrast, therapists and trained raters were much more consistent with treatment experts. In terms of practical significance, youth and caregivers each had a 97% estimated probability of indicating that a typical treatment component had occurred. By comparison, the probability was 31%, 19%, and 26% for therapists, trained raters, and treatment experts, respectively.

Conclusions—Youth and caregivers were highly inaccurate relative to treatment experts; whereas, therapists and trained raters were generally consistent with treatment experts. The implications of these findings for therapist adherence measurement are considered.

Keywords

Therapist adherence; rater method; contingency management; substance use

The measurement of therapist adherence to evidence-based practices has become a central feature of efforts to improve implementation of these practices in substance abuse and mental health services (Schoenwald, Garland, Chapman et al., 2011). This, in part, is attributable to empirical findings linking therapist adherence and desirable clinical outcomes

Correspondence concerning this article should be addressed to Jason E. Chapman, Ph.D., Medical University of South Carolina, Family Services Research Center, 67 President St., STE MC406, PO Box 2508861, Charleston, SC 29425. chapmaja@musc.edu.

(e.g., Hogue, et al., 2008; Sexton & Turner, 2010). Broadly defined, therapist adherence reflects the degree to which a therapist implements specific treatment components as intended (Hogue, Liddle, & Rowe, 1996).

Unfortunately, there is little consensus on the optimal approaches for measuring therapist adherence in research and practice settings. Observational coding with trained raters, the most common rating method in research trials, provides highly objective information (Hogue et al., 1996; Mowbray, Holter, Teague, & Bybee, 2003). However, this approach can require considerable time and expense. Therapist self-ratings also have been employed (e.g., Weersing, Weisz, & Donenberg, 2002) and are less costly than observational coding methods. However, findings indicate that therapists tend to over-report their use of treatment techniques (e.g., Carroll, Nich, & Rounsaville, 1998; Martino, Ball, Nich, Frankforter, & Carroll, 2009). Clients have also been used as raters of therapist adherence (see Schoenwald et al., 2011), serving to increase the feasibility, and reduce the cost, of measurement. However, these are "untrained raters" and researchers have yet to establish whether clients can accurately rate the occurrence of specific therapeutic techniques.

A few studies of substance abuse treatments have examined the relative performance of different rating methods, and have found low correspondence between therapist and observer ratings (Carroll et al., 1998; Martino et al., 2009). Similarly, Hurlburt, Garland, Nguyen, and Brookman-Frazee (2010) found that therapists and observational coders differed significantly in their perceptions of what occurred during community-based mental health treatment. Neither study, however, evaluated clients as raters of therapist adherence.

An on-going clinical trial provided a unique opportunity to compare the performance of four adherence rater types: clients (youth and caregivers), therapists, trained raters, and treatment experts. The trial is evaluating Contingency Management combined with a sexual risk reduction intervention (CM/SRR) for reducing substance use and risky sexual behavior among youth in juvenile drug court (JDC). Given budgetary constraints of the trial, youth and caregivers were initially selected as the raters of therapist adherence to the intervention protocol. Therapist ratings were later obtained to provide a check on these reports. Preliminary analyses found that therapists reported much lower levels of CM/SRR implementation relative to youth and caregivers (Chapman, McCart, Sheidow, & Letourneau, 2010). The findings, however, were inconclusive without objective ratings for the same cases. Thus, funding was obtained to supplement these ratings with observational ratings provided by trained raters and CM/SRR treatment experts. The goal of the current study was to evaluate the accuracy of youth and caregivers, therapists, and trained raters on ratings of therapist adherence, with "benchmark" ratings provided by treatment experts.

Method

Parent Study Design and Overview

The ongoing parent project uses a randomized design to evaluate the effectiveness of the CM/SRR intervention for JDC-involved youth and their caregivers. Two JDCs in the southeastern US refer youth. Newly referred youth are recruited and randomized to receive CM/SRR or usual services provided by community treatment facilities. The present study focused on therapist adherence to CM, excluding SRR protocol adherence and usual services.

CM Protocol

The evidence-based CM protocol (Henggeler et al., 2012) comes from research focused on implementing CM in community-based settings (e.g., Henggeler et al., 2006; Henggeler et al., 2008). The CM protocol is conducted as a family-based approach, and the components

are: Antecedent-Behavior-Consequence (ABC) Assessment, Point-and-Level System, Drug Testing, Self-Management Planning, and Drug Refusal Skills Training.

Therapist Training and Supervision

Three protocol therapists were hired by the study team and trained to deliver the CM/SRR intervention. Professionally, the therapists all held a Master's degree in Social Work, and they had 1–3 years of post-graduate clinical experience with adolescents and their families. An initial 12 hours of training oriented the therapists to program philosophy and intervention methods. The second author, a clinical psychologist, reviewed 2–3 complete session tapes per therapist per week and provided weekly individual supervision lasting one hour. Supervision aimed to promote adherence to the treatment protocol and develop solutions to difficult clinical problems.

CM Therapist Adherence Measure (CM-TAM)

The CM-TAM was initially developed for use in a JDC study (Henggeler et al., 2006). Subsequently, Many Facet Rasch Models (MFRM; Linacre, 1994) using data from three CM studies found that the instrument performed well, with items targeted to the sample, high item reliability, and stable performance over time (Chapman, Sheidow, Henggeler, Halliday-Boykins, & Cunningham, 2008). Improvements made based on the findings included: revising the content of some items, adding items to target a wider range of therapist adherence, and revising the 5-point rating scale. The revised CM-TAM has 34 items rated on a 4-point scale (Never, Rarely, Often, Almost Always) and was implemented in a recent study (McCart, Henggeler, Chapman, & Cunningham, 2012). The CM-TAM items are presented in Table 1. As part of the parent study, researchers administered the CM-TAM to youth and caregivers on a monthly basis beginning one month after consent and continuing through the end of treatment.

Current Study Overview

From the parent study, CM-TAM reports were obtained from 27 families (i.e., *youth and caregivers*). The three *therapists* used their detailed treatment notes and in-session checklists to provide retrospective, session-by-session ratings using the CM-TAM. Five *trained raters* and two *treatment experts* (the second and fourth authors) independently listened to audiotapes of treatment sessions and provided ratings on the CM-TAM. Because therapists, trained raters, and treatment experts rated individual sessions, their ratings were dichotomous for each item (i.e., did not occur v. did occur). All procedures were approved by the IRB at MUSC.

Youth and Caregiver Participants

The youth participants for the present study averaged 16 years of age (range = 14-17), 52% were White, 41% were Black, 26% were of Hispanic ethnicity, and 67% were male. Of these participants, 93% lived with a biological or adoptive parent, and 30% of the households received some form of financial assistance.

Rater Training

Rater training matched procedures used in previous studies (Sheidow, Donohue, Hill, Henggeler, & Ford, 2008). Five bachelors- to masters-level research assistants with no prior experience delivering CM completed six hours of training, conducted by the treatment experts. The training included detailed review of the CM-TAM and group practice ratings using taped examples. Next, the raters independently rated session tapes previously rated by the experts. Raters were "certified" after their ratings achieved acceptable consistency (

80% agreement) with the experts' ratings. Monthly booster sessions were held to monitor and address rater drift.

Rating Plan

The rating plan was designed to provide sufficient overlap across raters for the comparisons of interest while managing the overall rating burden (Linacre, 1994). Youth and caregivers completed 82 monthly CM-TAMs that referenced 148 sessions. Of these, 82% were completed by both respondents, and the other 18% were completed by the youth (25%) or caregiver (75%). Therapist ratings were available for 140 (95%) of the 148 sessions. Each session recording was rated by one (16%), two (57%), three (17%), or four (10%) trained raters. The two treatment experts rated a total of 100 sessions, overlapping on 18% of the recordings.

Statistical Analysis

The accuracy of the different types of raters was evaluated using Many-Facet Rasch Models (MFRMs; Linacre, 1994). The Rasch model is a special case of an Item Response Theory measurement model. It is a probabilistic model where, for example, the probability of a CM component occurring is the net result of (a) the difficulty of the CM component (i.e., the item) and (b) the therapist's level of CM adherence (i.e., the person; Wright & Mok, 2000). Thus, a highly adherent therapist has a high probability of implementing an easy CM component, and vice versa. The MFRM extends the basic model (i.e., items, persons) to evaluate and adjust for additional model "facets." More specifically, this model was developed to accommodate and evaluate rater data (e.g., Myford & Wolfe, 2003). As such, the MFRM is ideally suited to the present research where the primary focus is comparing the performance of different types of raters. The results of the MFRM provide information for each of the facets, including logit-based scores, *SEs*, and fit statistics. The results also permit tests for differences between rater types using their Rasch-based scores and *SEs* (Luppescu, 1995). The models were estimated using FACETS software (Linacre, 2011).

Two models were performed. The first model focused on the accuracy of youth and caregiver reports. Youth and caregivers provided monthly ratings – referencing 1 session in each report. Therapists, trained raters, and treatment experts provided ratings of individual sessions. Therefore, their ratings had to be aggregated to a monthly format. Across raters, the monthly format item ratings were dichotomized (i.e., did not occur, did occur). The resulting model included facets for month, family, therapist, rater type, and items. The second model more closely evaluated the accuracy of therapists and trained raters using their session-by-session ratings. For this model, the first facet became "session number" and the "rater type" facet did not include youth and caregivers.

Results

Performance of Youth and Caregiver Raters (Monthly Data)

The distribution for each facet is illustrated in Figure 1. The left-most column provides the scores for the elements of the facets. Consistent with the treatment protocol, the use of CM components was more common in earlier months of treatment. Families varied considerably in the level of CM they received, though this is likely due to only a partial course of treatment being rated. The three therapists were generally consistent in their use of CM, as expected due to intensive training and supervision. The CM-TAM items had a wide distribution and were well-targeted to the families (i.e., the family and item distributions overlap considerably).

For this study, the Rater Type facet is of greatest interest. The different rater types should be largely indistinguishable in rating style after accounting for the other facets in the model. The results, however, indicate otherwise. The location of "Expert" on the figure illustrates the position for the benchmark, "correct" rating style. Youth and caregivers were much more likely to endorse the occurrence of CM components than were experts. These differences were statistically significant, as evidenced by the non-overlapping 95% confidence intervals for youth [2, 41, 2, 91] and caregivers [2, 60, 3, 18] relative to experts [-1, 95, -1, 67]. In

for youth [2.41, 2.91] and caregivers [2.60, 3.18] relative to experts [-1.95, -1.67]. In contrast, therapist raters and trained raters were much more consistent with the treatment experts. However, the differences were statistically significant, with non-overlapping 95% CIs for therapist raters [-2.25, -2.09] and trained raters [-1.69, -1.45] relative to expert raters [-1.95, -1.67]. Trained raters were somewhat less likely to endorse CM components, and therapists were somewhat more likely to endorse CM components. Because raters are intended to provide accurate and highly consistent reports, Linacre (2011) recommends ~90% agreement in ratings. For this model, the agreement percentage was 57%.

To highlight the practical implications of these findings, we computed the probability of the different types of raters endorsing the occurrence of a specific CM component. For a typical component (e.g., Item 4, "The therapist conducted an ABC Assessment of my child's drug use"), youth and caregiver raters had a 97% probability of indicating that it had occurred. In contrast, therapists and trained raters had a 31% and 19% probability, respectively, of indicating that the same component had occurred. For the benchmark treatment expert raters, the probability was 26%. Thus, conclusions about the occurrence of CM components vary drastically, and in a practically meaningful way, between youth and caregivers and the other types of raters.

Performance of Therapist Raters and Trained Raters (Session-by-Session Data)

The second model used the session-by-session data to provide a more precise comparison of therapist raters and trained raters. The results are illustrated in Figure 2. The rater type facet indicates that the therapist raters and trained raters were again generally consistent with the treatment expert raters. However, as before, the differences that did exist reached statistical significance, with non-overlapping 95% CIs for therapist raters [-0.47, -0.35] and trained raters [0.31, 0.47] relative to expert raters [-0.08, 0.12]. Trained raters were somewhat less inclined to endorse the occurrence of CM components, and therapist raters were somewhat more inclined to endorse these components. The percentage of exact agreements for this model increased to 86%. For the same CM component (e.g., item 4), the probability of endorsement by therapists, trained raters, and treatment experts was 12%, 6%, and 9%.

Discussion

This study evaluated the accuracy of youth, caregivers, therapists, and trained raters relative to treatment experts on ratings of therapist adherence to CM for adolescent substance abuse. Youth and caregivers were consistent with each other; however, they reported much greater occurrence of CM components than did the treatment expert raters, trained raters, and therapist raters. Relative to youth and caregivers, therapists and trained raters were far more consistent with the treatment experts. However, these raters did differ from the experts, with therapists somewhat over-reporting and trained raters somewhat under-reporting the occurrence of CM. These findings suggest clients are less accurate in their assessment of what transpires in treatment sessions than therapists delivering the treatment, individuals trained to rate adherence, and treatment content experts. This finding is not surprising. Therapists and experts are trained in the intervention and treat multiple clients over time, thereby accruing experience in discerning and delivering treatment components. Similarly, trained raters are taught how to discern the occurrence of treatment components, and they rate multiple sessions from multiple therapists/clients over time. Clients, however, are not

trained in the intervention and typically have a single course of treatment. Thus, relative to the other raters, they likely lack the depth and breadth of training in and exposure to treatment components necessary for accurate ratings. Relative to treatment experts, therapists were more lenient and trained raters were more severe. Therapists, who are highly trained in the intervention but not the rating protocol, may have a lower threshold for indicating use of a treatment component. Trained raters, however, who are highly trained in the rating protocol but not the intervention, may have more difficulty detecting variation in the delivery of treatment components.

The practical significance of the observed differences is not trivial. On average, youth and caregivers indicated that 95% and 96% of the CM components had occurred. In contrast, therapists and trained raters reported 39% and 28%, and treatment experts reported 34%. Therefore, conclusions about therapist adherence vary dramatically across the different types of raters. Ultimately, the impact of such a difference depends on the types of decisions to be made using the scores. For the parent study, the scores are required to demonstrate that the intervention was delivered as intended in the experimental condition and not delivered in the comparison condition. Using the same descriptive statistics, to date, youth and caregivers in the usual services group have reported less occurrence of CM components, 68% and 62%, respectively. Similarly, a recent study demonstrated between-group differences using youth and caregiver reports on the CM-TAM (McCart et al., 2012). This suggests that youth and caregiver reports could be sufficient for a gross distinction between "receiving CM" and "not receiving CM." However, the level of over-reporting of CM components by youth and caregivers raises question with such a conclusion. For instance, caregivers in the usual services condition reported 2.2 times the level of CM use relative to trained raters rating the CM condition. Thus, it seems likely that more accurate reports would be required to improve therapist training and supervision, improve adherence to CM components, and evaluate the performance of therapists.

Several limitations of the study are important to note. First, although the primary focus was on monthly and session-by-session ratings for each family – yielding a larger sample of ratings for comparison – the overall sample of families was relatively small (N = 27). This warrants some caution when generalizing the findings. Next, to evaluate youth and caregiver raters, session-by-session ratings from the other raters were aggregated, and the responses for all raters were dichotomized. An alternative strategy could yield different findings (e.g., a 4-point rating scale, dichotomizing youth and caregiver ratings with a different threshold for "did not occur" versus "did occur"). However, one alternative was tested as a follow-up (i.e., recoding the maximum rating as "did occur" and all other ratings as "did not occur"), and the results were essentially identical. Also related to this, the extent to which youth and caregiver accuracy could improve if providing ratings of a single session (e.g., administration after each session), rather than monthly ratings, is currently unknown. Youth and caregiver over-endorsement of CM components could be attributable to inaccurate recall of the previous month's sessions. Further, therapist raters were found to provide generally accurate ratings, a finding that is encouraging with respect to feasibility. However, therapists' awareness that treatment supervisors routinely reviewed their session tapes may have contributed to greater vigilance and accuracy in reporting than would be found otherwise (e.g., Azrin et al., 2001). Likewise, the specific therapeutic components of CM might be easier for therapists to rate than the components of other interventions. Further, there were no consequences for therapists' reports (e.g., remediation, pay-for-performance), controlling potential demand characteristics. Finally, measurement models for therapist adherence ratings will ultimately need to be developed to accommodate interventionspecific features such as time-varying applicability of therapeutic components over the course of treatment (e.g., components intended to occur on a single occasion versus components intended to occur on an as-needed basis).

In conclusion, this is the first study to evaluate simultaneously ratings of therapist adherence provided by a full array of raters, ranging from youth and caregivers to treatment experts. The findings indicate that the ratings provided by youth and caregivers are far less accurate than those provided by therapists and trained raters.

Acknowledgments

This research was supported by grant R01DA025880 from the National Institute on Drug Abuse awarded to the third author and a sub-award to the first and second authors from National Institute on Mental Health grant P30MH074678 (John A. Landsverk, Ph.D., PI). The authors would like to thank Sonja K. Schoenwald, Ph.D. for her guidance on this project, including her comments and suggestions for this article.

References

- Azrin NH, Donohue B, Teichner GA, Crum T, Howell J, DeCato LA. A controlled evaluation and description of individual-cognitive problem solving and family-behavior therapies in duallydiagnosed conduct-disordered and substance-dependent youth. Journal of Child and Adolescent Substance Abuse. 2001; 11:1–43.
- Carroll KM, Nich C, Rounsaville BJ. Utility of therapist session checklists to monitor delivery of coping skills treatment for cocaine abusers. Psychotherapy Research. 1998; 8:307–320.
- Chapman, JE.; McCart, MR.; Sheidow, AJ.; Letourneau, EJ. The use of Rasch and Many-Facet Rasch models to compare untrained and partially-trained raters in the measurement of therapist adherence. Paper presented at the International Conference on Outcomes Measurement; Bethesda, Maryland. 2010.
- Chapman JE, Sheidow AJ, Henggeler SW, Halliday-Boykins CA, Cunningham PB. Developing a measure of therapist adherence to contingency management: An application of the many-facet Rasch model. Journal of Child and Adolescent Substance Abuse. 2008; 17:47–68. [PubMed: 20871744]
- Henggeler SW, Chapman JE, Rowland MD, Halliday-Boykins CA, Randall J, Shackelford J, Schoenwald SK. Statewide adoption and initial implementation of contingency management for substance abusing adolescents. Journal of Consulting and Clinical Psychology. 2008; 76:556–567. [PubMed: 18665685]
- Henggeler, SW.; Cunningham, PB.; Rowland, MD.; Schoenwald, SK.; Swenson, CC.; Sheidow, AJ.; Randall, J. Contingency management for adolescent substance abuse: A practitioner's guide. New York: Guilford Press; 2012.
- Henggeler SW, Halliday-Boykins CA, Cunningham PB, Randall J, Shapiro SB, Chapman JE. Juvenile drug court: Enhancing outcomes by integrating evidence-based treatments. Journal of Consulting and Clinical Psychology. 2006; 74:42–54. [PubMed: 16551142]
- Hogue A, Henderson CE, Dauber S, Barajas PS, Fried A, Liddle HA. 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]
- Hogue A, Liddle HA, Rowe C. Treatment adherence process research in family therapy: A rationale and some practical guidelines. Psychotherapy. 1996; 33:332–345.
- Hurlburt MS, Garland AF, Nguyen K, Brookman-Frazee L. Child and family therapy process: Concordance of therapist and observational perspectives. Administration and Policy in Mental Health and Mental Health Services Research. 2010; 37:230–244. [PubMed: 19902347]
- Linacre, JM. Many-Facet Rasch Measurement. Chicago: MESA Press; 1994.
- Linacre, JM. Facets computer program for many-facet Rasch measurement (version 3.68.1) [Computer software and manual]. Beaverton, Oregon: Winsteps.com; 2011.
- Luppescu S. Comparing measures: Scatterplots. Rasch Measurement Transactions. 1995; 9:410.
- Martino S, Ball S, Nich C, Frankforter TL, Carroll KM. Correspondence of motivational enhancement treatment integrity ratings among therapists, supervisors, and observers. Psychotherapy Research. 2009; 19:181–193. [PubMed: 19396649]

- McCart MR, Henggeler SW, Chapman JE, Cunningham PB. System-level effects of integrating a promising treatment into juvenile drug courts. Journal of Substance Abuse Treatment. 2012; 43:231–243. [PubMed: 22154039]
- Mowbray CT, Holter MC, Teague GB, Bybee D. Fidelity criteria: Development, measurement, & validation. American Journal of Evaluation. 2003; 24:315–340.
- Myford CM, Wolfe EW. Detecting and measuring rater effects using Many-Facet Rasch Measurement: Part I. Journal of Applied Measurement. 2003; 4:386–422. [PubMed: 14523257]
- Sexton T, Turner CW. The effectiveness of functional family therapy for youth with behavioral problems in a community practice setting. Journal of Family Psychology. 2010; 24:339–348. [PubMed: 20545407]
- Sheidow AJ, Donohue BC, Hill HH, Henggeler SW, Ford JD. Development of an audio-tape review system for supporting adherence to an evidence-based treatment. Professional Psychology: Research and Practice. 2008; 39:553–560.
- Schoenwald SK, Garland AF, Chapman JE, Frazier SL, Sheidow AJ, Southam-Gerow MA. Toward the effective and efficient measurement of implementation fidelity. Administration and Policy in Mental Health and Mental Health Services Research. 2011; 38:32–43. [PubMed: 20957425]
- Weersing VR, Weisz JR, Donenberg GR. Development of the therapy procedures checklist: A therapist-report measure of technique use in child and adolescent treatment. Journal of Clinical Child Psychology. 2002; 31:168–180.
- Wright BD, Mok M. Rasch models overview. Journal of Applied Measurement. 2000; 1:83–106. [PubMed: 12023559]

SCORE	MONTH More CM	FAMILY Received More CM	THERAPIST Used More CM	RATER TYPE Endorsed More CM	ITEMS More Difficult CM Component
3 		 	+ 	 CAREGIVER YOUTH 	+
 2 	4	* 	- 	- 	
 1	2 3	 * * * * * *	 	- 	30 17 34
 0 	1 5	* * * * * * * *	 # # #		7 13 2 33 1 25 31 32 18 26 28 29 11 24 27 23
 -1 	6	* * * * *	 	 	9 12 20 4 5 6 10 21 3 10 22
 -2 		 * * *	 	THERAPIST EXPERT TRAINED 	19 22 8 14 16
 -3 	9		 ·	 + 	 15
 -4					
+	Less CM	Received	Used Less CM	Endorsed Less CM	Less Difficult CM Component

Figure 1.

Illustration of the results of the Many-Facet Rasch Model using monthly data and comparing the performance of youth and caregiver raters, therapist raters, and trained raters to treatment expert raters. The vertical positions depict the score for each month, family, therapist, rater type, and item.

SCORE	SESSION More CM	FAMILY Received More CM	THERAPIST Used More CM	RATER TYPE Endorsed More CM	ITEMS More Difficult CM Component
5 		 	 	 	 30 34
4 		 	 		 2 17 33 31 32 1 13 25 7
3 			 	 	26 29 27 28 11 18 5 23 24 4 6 9 12
2 		* 	 	 	20 10 21 3
	10 16 3 4 13 14 15 2 17 7 9 11 12 6	 * * * * * *	 		22
	1 18 5 24 26 25 20	* * * * * * * *	# # 	EXPERT TRAINED	14
 _1 	19 22 21 23 	* * * * * * *	# 	 	 15
 -2 	 29 	 	 	 	
 -3 	30 31 32 	 			
-4 +	 +	 +	 +	 +	ı +
	Less CM	Received Less CM	Used Less CM	Endorsed Less CM	Less Difficult CM Component

Figure 2.

Illustration of the results of the Many-Facet Rasch Model using session-by-session data and comparing the performance of therapist raters and trained raters to treatment expert raters. The vertical positions depict the score for each session, family, therapist, rater type, and item.

Table 1

Caregiver-Reported Contingency Management Therapist Adherence Measure (CM-TAM) Items

#	Item Text
1.	The therapist explained what an ABC Assessment is and why it is used.
2.	The therapist helped my family think of relapses as "learning opportunities" rather than "failures."
3.	The therapist talked with us about situations that might cause or "trigger" my child to use drugs.
4.	The therapist conducted an ABC Assessment of my child's drug use.
5.	In the ABC Assessment, the therapist made sure that "Behaviors" included things my child did to obtain drugs.
6.	In the ABC Assessment, the therapist made sure that "Consequences" included both the positive and negative consequences of my child's drug use.
7.	The therapist made sure that my child and I completed an ABC Assessment on our own.
8.	The therapist used a contract with us that focused on stopping drug use in exchange for rewards.
9.	The therapist identified my child's Most Valued Privilege (that is, a privilege that my child would work really hard to get).
10.	The therapist worked with us to identify rewards for my child.
11.	The therapist made sure that rewards included a balance between vouchers (such as gift cards) that the therapist provided and privileges that I provided.
12.	The therapist made sure that any rewards given to or withheld from my child would motivate him/her to stop using drugs.
13.	The therapist helped me focus on rewarding abstinence rather than focusing on punishment for drug use.
14.	The therapist used the checkbook, or another similar system, to track my child's points.
15.	The therapist tested my child for drug use using a drug screen.
16.	The therapist made sure that I tested my child for drug use using a drug screen.
17.	The therapist made sure that an ABC Assessment was completed after each drug screen, regardless of whether the result was dirty or clean.
18.	The therapist made sure that I provided or withheld the Most Valued Privilege based on my child's drug screen results.
19.	The therapist made sure that positive or negative consequences were given based on the results of my child's drug screen.
20.	The therapist made sure the drug screens were administered frequently enough to catch drug use.
21.	The therapist made sure the drug screens were administered randomly (at times my child was not expecting it).
22.	The therapist made sure strategies were used to prevent my child from trying to alter the results of his/her drug screens.
23.	The therapist made sure that drug screens were administered at times when my child was more likely to use drugs.
24.	The therapist worked with us on a Self-Management Plan regarding my child's drug use.
25.	The therapist made sure that my child's Self-Management Plan was based on information we had provided in ABC Assessments.
26.	The therapist helped us think of ways to rearrange our environment to help my child avoid situations that might cause or "trigger" drug use.
27.	The therapist helped my child think of ways to avoid situations that might cause or "trigger" drug use.
28.	The therapist made sure that the drug avoidance strategies were best suited for my child's circumstances, peer group, and personality.
29.	The therapist helped my child think of the costs and benefits of different drug avoidance strategies.
30.	The therapist helped my child practice drug avoidance strategies.
31.	The therapist helped my child think of ways to refuse drugs.
32.	The therapist made sure that the drug refusal strategies were best suited for my child's circumstances, peer group, and personality.
33.	The therapist helped my child think of the costs and benefits of different drug refusal strategies.
34.	The therapist helped my child practice drug refusal strategies.