Effects of Routine Feedback to Clinicians on Mental Health Outcomes of Youths: Results of a Randomized Trial

Leonard Bickman, Ph.D.
Susan Douglas Kelley, Ph.D.
Carolyn Breda, Ph.D.
Ana Regina de Andrade, Ph.D.
Manuel Riemer, Ph.D.

Objective: A randomized cluster controlled trial tested the hypothesis that weekly feedback to clinicians would improve the effectiveness of home-based mental health treatment received by youths in community settings. *Methods:* Youths, caregivers, and clinicians at 28 sites in ten states completed assessments of the youths' symptoms and functioning every other week. Clinicians at 13 sites were provided with weekly feedback about the assessments, and clinicians at 15 sites received feedback every 90 days. Data were collected from June 1, 2006, through December 31, 2008. Intent-to-treat analyses were conducted with hierarchical linear modeling of data provided by youths, caregivers, and clinicians. *Results:* Assessments by youths, caregivers, and clinicians indicated that youths (N=173) treated at sites where clinicians could receive weekly feedback improved faster than youths (N=167) treated at sites where clinicians did not receive weekly feedback. A dose-response analysis showed even stronger effects when clinicians viewed more feedback reports. *Conclusions:* Routine measurement and feedback can be used to improve outcomes for youths who receive typical home-based services in the community. (*Psychiatric Services* 62:1423–1429, 2011)

Thousands of studies unrelated to behavioral health show that without measuring performance and providing feedback, improvement is minimal (1). In the behavioral health field, greater use of outcome monitoring and feedback has been recommended for all practices (2–4).

A measurement feedback system is one tool for providing outcome monitoring and feedback in a clinical setting (1). It provides systematic and frequent measurement of treatment progress and processes within a continuous quality improvement framework. It is designed to provide feedback that enhances clinical decision making, improves accountability, drives program planning, and informs treatment effectiveness (5–7). To effectively and continually improve performance, the feedback must be accurate and clinically useful (1,6).

Research with adults supports the use of systematic feedback to clinicians (8,9). For example, clinicians' ability to detect worsening of adult clients' symptoms is improved through regular and systematic outcome monitoring and feedback (10,11). Worthen and Lambert (12) have suggested that feedback influences clinical outcomes by providing clinicians with information they may have unintentionally overlooked or underemphasized and by identifying problems within domains that could jeopardize progress. Feedback can assess actual client change, enhance therapeutic alliance, produce more accurate case conceptualizations, and foster richer discussions of potential change in treatment plans (13).

However, there are no clinical trials of the effectiveness of measurement feedback systems in improving outcomes for youths. Published studies have focused on establishing the accuracy of warning systems that can be used to predict youth treatment failure on the basis of outcome measures scored with computerized algorithms (14,15). One study found that incorporating data from multiple reporters, for example, youths and caregivers, is most sensitive in identifying youths at risk for treatment failure (14).

This study used a cluster randomized experimental design to test the hypothesis that weekly feedback to clinicians improves the effectiveness of mental health treatment of youths living in community settings. Feedback is provided by Contextualized Feedback Systems (CFS), which include a psychometrically sound and

Dr. Bickman, Dr. Kelley, Dr. Breda, and Dr. de Andrade are affiliated with the Center for Evaluation and Program Improvement, Vanderbilt University, Peabody 151, 230 Appleton Pl., Nashville, TN 37203-5721 (e-mail: leonard.bickman@vanderbilt.edu). Dr. Riemer is with the Department of Psychology, Wilfrid Laurier University, Waterloo, Ontario, Canada.

clinically useful battery of very brief measures that promotes overall practice improvement through frequent and comprehensive assessments (16). CFS is based on a theoretical synthesis of perspectives from social cognitive psychology, organizational theory, and management that explains the mediating and moderating structure of feedback interventions that can change clinician behavior (17–19). An earlier version of CFS called Contextualized Feedback Intervention and Training was used in this study.

Methods

Design and procedures

The study design and procedures were selected to provide the optimal balance between the requirements for scientific rigor and the need to test the intervention under real-world conditions. Sites affiliated with the Providence Service Corporation (PSC), a private, for-profit, behavioral health organization, participated in the study. PSC provides mainly services to youths and their families in their homes. As a highly decentralized organization, its services are not uniform across sites. Specific type of treatment is not prescribed. Clinicians report using various therapeutic approaches, including cognitive-behavioral, integrative-eclectic, behavioral, family systems, and play therapy.

All PSC sites in the study agreed to adopt CFS as part of the organization's ongoing continuous quality improvement initiative with the intention of learning whether its use was financially viable and organizationally feasible. Forty-nine sites expressed interest in participating in the randomized clinical trial of CFS. They were randomly assigned by the research study's data manager to an experimental or control group. The control group clinicians could receive feedback on a client every 90 days. Subsequently, 21 sites (11 experimental and ten control) dropped out of the study. Our initial plan specified a 2×2 factorial design that included two conditions—feedback and the provision of three Web-based modules on common factors (therapeutic alliance, expectancies about counseling, and collaborative treatment planning). However, only one-third (N=31) of

the clinicians accessed the training modules before their first client. Thus, we considered this module condition to be an implementation failure and analyzed data by the feedback condition only. There were no statistical differences in reasons for attrition between experimental and control groups.

Individual support was available thereafter by phone or e-mail. Initial training on how to integrate feedback into practice was provided through on-site workshops. Ongoing training was provided through regularly scheduled (at least monthly) group teleconferences. CFS was introduced to sites as a continuous quality improvement initiative. Thus clinicians were expected and encouraged to participate. However, some clinicians did not participate and some participated with only some clients. The research project provided the clinicians at the 28 sites with a youth and caregiver brochure to give to clients that described CFS. It also trained clinicians in how to explain CFS to clients. Training was segregated by study group assignment so that its content aligned with either the weekly or the 90-day feedback control condition.

At the close of a treatment session, the youth, caregiver, and clinician completed paper questionnaires. They all placed their completed forms in an envelope that was sealed by the clinician. The clinician delivered the envelope to a project-trained assistant at each site, who entered the data. Feedback reports were available as soon as data were entered into the system. The researchers did not collect or enter data and received a limited data set for analyses that included dates of treatment but no other personal identifiers. Data collection extended from June 1, 2006, through December 31, 2008.

All youths aged 11 to 18 who entered home-based services after CFS implementation began at each site were eligible to participate. The study was approved by the Vanderbilt University Institutional Review Board with a waiver of informed consent.

Feedback intervention

Clinicians in the experimental group received weekly feedback plus cumulative feedback every 90 days after a

youth was enrolled in CFS. The weekly reports were available a median of nine days (mean±SD=12.3±22.3) after the end of the session. Nearly half (46%) of these reports were available within a week or less.

Clinicians in the control group received only the cumulative, 90-day feedback. Because youths remained in CFS about four months (mean= 3.8±3.1; median=3.3), many would have been discharged before their first 90-day report became available. We considered the 90-day group to be a no-feedback control group.

Feedback was automatically generated by CFS in the form of computer screens that compared and summarized measures completed by the youth, caregiver, and clinician at previous sessions. Examples of feedback included mean scores and alerts if the youth's symptoms ranked in the top 25th percentile in severity. Indicators of whether change from one measurement instance to the next met criteria for reliable change and trend graphs for change over multiple measurement points were also provided. Feedback viewing was tracked in the system whenever clinicians clicked the "radio button" at the bottom of the main feedback Web page indicating if they agreed with the report. A detailed description of the current version of the feedback system, with screen shots, is available at cfsystemsonline.com.

Measures

Youths' symptom severity and functioning were assessed by the Symptoms and Functioning Severity Scale (SFSS), part of the Peabody Treatment Progress Battery (16,20). The SFSS was completed by the youth, the caregiver, and the clinician to provide data from multiple perspectives. Cronbach's alphas (21) for the youth, caregiver, and clinician forms are .92, .94, and .93, respectively. Correlations with the Youth Self-Report, the Strength and Difficulties Questionnaire, and the Youth Outcomes Questionnaire—other measures of youths' symptomatology—range from .71 to .89, depending on respondent type, indicating good convergent validity (22–24).

The SFSS assesses change over time in closely timed repeated measurements. Version 1 of the SFSS con-

sists of 32 items that rate how frequently within the last two weeks the youth experienced emotions or exhibited behaviors linked to typical mental health disorders among youths, including attention-deficit hyperactivity disorder, conduct or oppositional defiant disorder depression, and anxiety. Frequency is rated 1, never; 2, hardly ever; 3, sometimes; 4, often; or 5, very often. A total severity scale score is created by a simple average of ratings for each youth if at least 85% of items are completed.

Statistical analyses

The main hypothesis tested was whether youths in the experimental group, whose clinicians could receive weekly feedback reports, improved faster than youths in the control group. This intention-to-treat analysis was repeated for data provided by each of the three respondents—youths, caregivers, and clinicians.

To test the hypothesis, we used hierarchical, longitudinal, slopes-as-outcome models that used random coefficients. In our hierarchical linear models (HLM), repeated measures were nested within participants, youths were nested within clinicians, and clinicians were nested within sites (25–28). We estimated three HLMs, one for each respondent type, to determine whether the feedback intervention had an effect on individuals' trajectories of outcome and whether there might be any differential effect by respondent type. We used SAS, Proc Mixed V9.12, to estimate the model by using restricted maximum likelihood (RML) estimation. HLMs, which include fixed effects and random intercepts for youth, clinician, and site levels, allowed for an exchangeable correlation structure at each level (29,30). RML estimation is recommended for multilevel models if repeated measures are not equally spaced (25). HLMs offer important advantages over older models (31–33), such as better handling of missing values and unequal time intervals between and within participant responses. Repeated measurements also increase the statistical power, describe the shape of change over time, and avoid the psychometric problems associated with changes in scores before and after an intervention.

Results

Sample

Twenty-eight sites in ten states were included in the analyses, 13 in the experimental group and 15 in the control group. Information about services and clientele at sites that were not included in the evaluation was not available. However, an organizational survey provided data on 24 of the 28 evaluation sites and 107 nonevaluation sites. The sites did not differ significantly on number, years employed, highest degree, or degree specialty of clinicians.

Table 1 shows background characteristics of youths in the experimental and control groups. Tables 2 and 3

Table 1

Baseline characteristics of 340 youths enrolled in Contextualized Feedback Systems (CFS), by treatment site[a]

Characteristic	Control (N=15)		Experimental (N=13)		Test statistic	df	p	Adjusted p
	N	%	N	%				
N of youths	167	49	173	51				
Gender					$\chi^2=2.02$	1	.156	.800
Female	87	54	79	46				
Male	75	46	93	54				
Ethnicity					$\chi^2=.41$	1	.521	1.00
Hispanic or Latino	14	10	18	13				
Not Hispanic or Latino	120	90	121	87				
Race					$\chi^2=16.39$[b]	1	.001	.01
White	97	68	71	45				
Nonwhite	46	32	88	55				
American Indian or Alaska Native	3	2	1	1				
Asian	0	0	2	1				
Black or African American	24	17	52	33				
More than one race	16	11	23	14				
Other	3	2	10	6				
Age (mean±SD years)	15.03±1.83		14.57±1.80		t=2.39	337	.018	.190
Baseline SFSS (mean±SD score)[d]								
Youth	2.34±.67		2.38±.72		t=.53	338	.596	1.00
Caregiver	2.52±.74		2.60±.78		t=.91	306	.365	.970
Clinician	2.25±.58		2.42±.52		t=2.73	292	.007	.130
Number of sessions (mean±SD)	11.36±9.78		10.68±8.54		F=.04[c]	1, 347	.834	1.00
Number of weeks in CFS (mean±SD)	18.07±15.53		14.91±11.23		F=1.53[c]	1, 347	.216	.370

[a] Clinicians at control sites could access feedback every 90 days; clinicians at experimental sites could access feedback weekly.
[b] Comparison was between categories of white and nonwhite only.
[c] Kruskal-Wallis test
[d] Possible scores for the Symptoms and Functioning Severity Scale (SFSS) range from 1 to 5, with higher scores indicating greater severity.

show characteristics of caregivers and clinicians.

The 340 youths who completed the SFSS at least once comprised the analytical sample used in the HLMs of youth reports of outcome. Multiple group testing is controlled with a bootstrap alpha on the basis of 100,000 resamples with replacement. After adjustment of the p values, we found no statistically significant differences at baseline between the control and the experimental groups with one exception. The experimental group had more black and fewer white youths (Table 1) and caregivers (Table 2) than the control group (p<.05). There were no significant differences in any clinician characteristic (Table 3).

Youths participated in the study for a mean±SD of 16.5±13.6 weeks (median=14.5). A total of 3,775 research records were generated, each representing a week in which the youth received treatment and any CFS data were collected. The mean number of research records per youth was 11±9.2, indicating that CFS data were not collected every week. At least one measure in the battery was scheduled to be collected every week, but the SFSS was scheduled to be completed every two weeks. However, it was not always possible to adhere to the schedule. The SFSS was completed 4.2±3.3, 3.0±2.8, and 4.2±3.8 times by youths, caregivers, and clinicians, respectively.

Baseline ratings on the SFSS varied between types of respondents—caregivers' ratings were significantly higher (2.56±.76) than ratings by clinicians (2.34±.55) and youths (2.36±.55) (p<.002). Weak agreement (weighted kappas <.26) between caregivers and clinicians and between youths and caregivers supported our decision to treat data from each type of respondent as an independent test of outcome.

Outcomes

We estimated the individual growth trajectories of the total SFSS score as reported by youths, clinicians, and caregivers. We adjusted each model by youths' race, the only variable found to be imbalanced at baseline

Table 2

Baseline characteristics of 383 caregivers of youths, by treatment site[a]

Characteristic	Control (N=15)		Experimental (N=13)		Test statistic	df	p	Adjusted p
	N	%	N	%				
N	169	44	214	56				
Gender					χ^2=2.59	1	.12	.73
Female	112	83	129	90				
Male	23	17	15	10				
Ethnicity					χ^2=.03	1	.87	1.00
Hispanic or Latino	9	7	10	8				
Not Hispanic or Latino	121	93	124	92				
Race					χ^2=22.42[b]	1	.001	.01
White	106	80	78	55				
Nonwhite	27	20	62	44				
American Indian or Alaska Native	2	2	1	1				
Asian	1	1	2	2				
Black or African American	18	14	52	37				
More than one race	2	2	3	2				
Other	4	3	4	3				
Marital status					χ^2=4.38	1	.50	.74
Never married	19	14	25	18				
Married	64	48	52	37				
Living with someone as married	5	4	7	5				
Widowed	5	4	8	6				
Divorced	29	22	32	23				
Separated	10	8	16	11				
Education					χ^2=4.52	1	.34	1.00
No formal schooling	1	1	1	1				
Less than high school or GED	26	20	38	28				
High school or GED	84	66	73	54				
Bachelor's degree	14	11	15	11				
Postgraduate degree	3	2	7	5				
Living with youth					χ^2=.42	1	.94	1.00
Not living together	3	2	3	2				
Some of the time	3	2	4	3				
Most of the time	5	4	7	5				
All of the time	125	92	126	90				
Age (mean±SD years)	42.31±8.63		43.96±10.9		t=1.55	335	.12	1.00

[a] Clinicians at control sites could access feedback every 90 days; clinicians at experimental sites could access feedback weekly.
[b] Comparison was between categories of white and nonwhite only.

Baseline characteristics of 144 clinicians, by treatment site[a]

Characteristic	Control (N=15)		Feedback (N=13)		Test statistic	df	p	Adjusted p
	Treatment site							
	N	%	N	%				
N of clinicians	64	44	80	56				
Gender					χ^2=.93	1	.33	.92
Female	42	84	53	77				
Male	8	16	16	23				
Ethnicity					χ^2=1.60	1	.21	.87
Hispanic or Latino	1	2	5	7				
Not Hispanic or Latino	48	98	64	93				
Race					χ^2=5.61[b]	1	.07	.39
White	35	73	36	56				
Nonwhite	13	27	28	44				
American Indian or Alaska Native	1	2	0	—				
Asian	0	—	0	—				
Black or African American	12	25	28	42				
More than one race	0	—	0	—				
Other	0	—	0	—				
Education					χ^2=4.70	1	.19	1.00
M.A. or M.S.	33	69	49	75				
B.A., B.S., or some graduate work	14	29	13	20				
Other	1	2	3	5				
Age (mean±SD years)	36.76±11.3		35.75±9.93		t=.51	114	.61	1.00

[a] Clinicians at control group sites could access feedback every 90 days; clinicians at feedback sites could access feedback weekly.
[b] Comparison was between categories of white and nonwhite only.

between experimental and control groups.

Table 4 shows the results of the intent-to-treat analysis. The intercept is the average SFSS for the two reference groups (control group and nonwhite youths) at baseline (time=0), and the other estimated parameters are deviations from the intercept. Regardless of type of respondent, the estimated feedback coefficient was not statistically significant, indicating that the control and experimental groups had the same level of functioning and symptomatology upon starting CFS; that is, there were no initial group differences.

Clinicians and caregivers reported that symptoms at baseline were more severe among white youths than among nonwhite youths (p=.001 and p=.02, respectively). Youths did not report a race difference (p=.54). Youths and clinicians reported significant improvement in youths' outcome over time (effect size=.30 and .17, respectively, data not shown), but caregivers did not. All three groups of respondents reported that youths in the feedback group improved significantly faster than youths in the control group (p<.01). Feedback effect sizes were .18, .24, and .27 for youths, clinicians, and caregivers, respectively. All effect sizes used the HLM-estimated coefficients measured at the average length of stay in CFS.

The intention-to-treat analysis described above did not consider that one-third of the clinicians in the feedback group did not view any feedback. A separate HLM analysis of the experimental and control groups was conducted to take into account whether a report was viewed (data available upon request). For all types of respondents, youths whose clinicians viewed at least one feedback report improved faster than youths whose clinicians did not view any report (p<.02). When we examined the proportion of reports viewed in a dose-response analysis, effect sizes increased by 50% for youths, to .27, and by 66% for clinicians, to .40 (p<.001). The effect size did not increase for caregivers.

Discussion

This is the first randomized controlled trial to examine the effects of feedback to clinicians on youths' clinical improvement. We found that regardless of type of respondent, youths whose clinician had access to weekly feedback improved faster than youths whose clinician did not.

The effect sizes for CFS were modest. However, to put them in context, they were about the same size as those found for comparisons of empirically supported treatments (ESTs) and treatment as usual in community settings (5). Moreover, it is encouraging to find an effect of CFS, consistent across three types of respondents, when used with treatments of unconfirmed effectiveness as usually found in community settings. CFS may be of even greater benefit to clinicians who use established ESTs. Moreover, because some clinicians resist using ESTs because they think they can harm the therapeutic relationship (34), CFS can help clinicians monitor this important relationship and adjust their approach accordingly. Combining feedback with established ESTs, a design we are now testing with a new software package that combines Functional Family Therapy with CFS (35), may optimize both approaches.

Table 4

Hierarchical longitudinal models of total scores for the SFSS, by respondent[a]

Parameter	Youths Estimated coefficient	SE	Clinicians Estimated coefficient	SE	Caregivers Estimated coefficient	SE
Fixed effects						
Intercept	2.30***	.09	2.20***	.08	2.39***	.11
Feedback group (reference: control group)[b]	.02	.10	.10	.10	.01	.13
Youth white (reference: nonwhite)[c]	−.05	.08	.18**	.06	.21*	.09
Slope (weeks)	−.01***	.002	−.005***	.001	−.003	.002
FB × slope[d]	−.01**	.002	−.01***	.002	−.01***	.003
Variance component						
Between youths	.29***	.03	.07***	.01	.31***	.04
Between clinicians	.02	.01	.13***	.03	.01	.01
Between sites	.03	.02	.01	.02	.05	.03
Within youths	.15***	.01	.10***	.004	.17***	.01
Restricted maximum likelihood estimation −2 restricted log likelihood	1,936		1,243		1,535	
Intraclass correlation coefficient (%)						
Between youths	60		23		57	
Between clinicians	4		42		2	
Between sites	5		3		9	
Within youths	31		32		31	

[a] Total number of Symptoms and Functioning Severity Scales (SFSS) submitted by youths (N=1,341), clinicians (N=1,291), and caregivers (N=935).

[b] Clinicians at experimental sites could access feedback weekly; clinicians at control sites could access feedback every 90 days.

[c] Including the slope-by-race interaction in the same models did not change the feedback effects (results not shown).

[d] Interaction term of feedback (FB) by slope; it estimates the additional rate of change that participants of the feedback group might have. A negative and significant FB × slope interaction coefficient is evidence that weekly feedback has a positive effect on youth's symptoms and functioning outcomes.

*p<.05
**p<.01
***p<.001

The dose-response analyses showed that the effect size was dramatically increased for two of the three types of respondent reports when the proportion of feedback reports viewed (dose) was considered. Increasing viewing of feedback may be a key to increasing its effects. However, because the dose analyses were correlational, it is possible that other variables accounted for the relationship between dose and changes in severity. For example, it could be that clinicians who viewed more reports were better clinicians.

Interviews with clinicians and supervisors as well as data from this study have been used to develop a third version of CFS that includes better clinician adherence monitoring capacity, easier implementation, and enhanced feedback reports. We anticipate that the latest version of CFS will improve implementation and produce even stronger effects than those reported here, but we need to learn more about how clinicians use feedback to plan and conduct treatment (36).

There were several limitations associated with conducting a large-scale, multisite field experiment of a measurement feedback system. Although sites were randomly assigned, clinicians within a site who were the most effective could have volunteered to use weekly feedback more often than clinicians in the control group, which received only cumulative feedback at 90 days. However, such a possibility would require that clinicians in the weekly feedback group were more highly motivated and that their self-perceptions of efficacy actually led them to be more effective. Such assumptions are not supported by the data.

It could also be possible that clients at sites where weekly feedback was provided were selected by their clinicians because they thought they were more likely to improve than clients at the control sites. Again, there is little support for this assumption, given that the only initial difference between the groups was race and not any variable that predicted improvement. Moreover, for this possibility to be true, we would need to assume that clinicians could predict who would improve faster, a fact not in evidence.

A study in which data are collected by the clinician involves a trade-off between the practical realities of a real-world evaluation and the stricter but entirely unaffordable protocols in which data are collected by the researcher. In addition, a study in which researchers collect the data would invalidate a study of a system that is designed for clinician data collection.

Significant attrition from the initial random assignments of sites could have biased the samples. However, except for race, which did not relate to improvement, there were no significant differences between the experimental and control groups. It is also possible that the groups differed in some unmeasured way that is correlated to how long they stayed in the study. Yet youths in the experimental and control groups attended an equivalent number of sessions and were enrolled for an equivalent length of time. Finally because we do not know the types of treatments provided at each site, it is possible that differential treatments could have influenced the outcome. However, none of the sites reported that they consistently used specific treatments.

Conclusions

Generalizability is a problem in most mental health research because it is very rare to conduct formal representative sampling of sites or services. The results of this study strictly apply to the sites studied and the services provided, in this case typical home-based care and the CFS intervention. However, a study of 28 sites across ten states that used a randomized

controlled trial is exceptional.

The results indicated that there was no significant heterogeneity among the sites and that attrition by sites did not appear to introduce bias in the data collected. This is the first study of youths in diverse and real-life community settings that has shown that mental health outcomes can be improved without necessarily introducing a more usual evidence-based treatment. It supports the use of measurement feedback systems in community clinical practice as an important approach to improving outcomes.

References

1. Bickman L: Why don't we have effective mental health services? Administration and Policy in Mental Health and Mental Health Services Research 35:437–439, 2008

2. APA Presidential Task Force on Evidence-Based Practice: Evidence-based practice in psychology. American Psychologist 61:271–285, 2006

3. Kazdin AE, Blase SL: Rebooting psychotherapy research and practice to reduce the burden of mental illness. Perspectives on Psychological Science 6:21–37, 2011

4. Newnham E, Page A: Bridging the gap between best evidence and best practice in mental health. Clinical Psychology Review 30:127–142, 2010

5. Garland A, Bickman L, Chorpita B: Change what? Identifying quality improvement targets by investigating usual mental health care. Administration and Policy in Mental Health and Mental Health Services Research 37:15–26, 2010

6. Kelley SD, Bickman L: Beyond outcomes monitoring: measurement feedback systems (MFS) in child and adolescent clinical practice. Current Opinion in Psychiatry 22:363–368, 2009

7. Chorpita B, Bernstein A, Daleiden E: Driving with roadmaps and dashboards: Using information resources to structure the decision models in service organizations. Administration and Policy in Mental Health and Mental Health Services Research 35:114–123, 2008

8. Knaup C, Koesters M, Schoefer D, et al: Effect of feedback of treatment outcome in specialist mental healthcare: meta-analysis. British Journal of Psychiatry 195:15–22, 2009

9. Reese R, Norsworthy L, Rowlands S: Does a continuous feedback system improve psychotherapy outcome? Psychotherapy: Theory, Research, Practice, Training 46:418–431, 2009

10. Lambert M: Yes, it is time for clinicians to routinely monitor treatment outcome; in The Heart and Soul of Change (2nd ed). Edited by Duncan BL, Miller SD, Wampold BE, et al. Washington, DC, American Psychological Association, 2010

11. Hatfield D, McCullough L, Plucinski A, et al: Do we know when our clients get worse? An investigation of therapists' ability to detect negative client change. Clinical Psychology and Psychotherapy 17:25–32, 2010

12. Worthen VE, Lambert MJ: Outcome oriented supervision: advantages of adding systematic client tracking to supportive consultations. Counseling and Psychotherapy Research 7:48–53, 2007

13. Hatfield DR, Ogles BM: The influence of outcome measures in assessing client change and treatment decisions. Journal of Clinical Psychology 62:325–338, 2006

14. Cannon J, Warren JS, Nelson PL, et al: Change trajectories for the youth outcome questionnaire self-report: identifying youth at risk for treatment failure. Journal of Clinical Child and Adolescent Psychology 39:289–301, 2010

15. Warren JS, Nelson PL, Mondragon SA, et al: Youth psychotherapy change trajectories and outcomes in usual care: community mental health versus managed care settings. Journal of Consulting and Clinical Psychology 78:144–155, 2010

16. Bickman L, Riemer M, Lambert EW, et al: Manual of the Peabody Treatment Progress Battery [electronic version]. Nashville, Tenn, Vanderbilt University, 2007

17. Bickman L, Riemer M, Breda C, et al: CFIT: a system to provide a continuous quality improvement infrastructure through organizational responsiveness, measurement, training, and feedback. Report on Emotional and Behavioral Disorders in Youth 6:86–94, 2006

18. Riemer M, Rosof-Williams J, Bickman L: Theories related to changing clinician practice. Child and Adolescent Psychiatric Clinics of North America 14:241–254, 2005

19. Riemer M, Bickman L: Using program theory to link social psychology and program evaluation; in Social Psychology and Program/Policy Evaluation. Edited by Mark MM, Donaldson SI, Campbell B. New York, Guilford, 2011

20. Bickman L, Athay M, Riemer M, et al: Manual of the Peabody Treatment Progress Battery, 2nd ed. Nashville, Tenn, Vanderbilt University, 2010

21. Cronbach LJ, Meehl PE: Construct validity in psychological tests; in Minnesota Studies in the Philosophy of Science, Vol 1: The Foundations of Science and the Concepts of Psychology and Psychoanalysis. Edited by Feigl H, Scriven M. Minneapolis, University of Minnesota Press, 1973

22. Achenbach TM: Integrative Guide for the 1991 CBCL/4-18, YSR, and TRF Profiles. Burlington, University of Vermont, Department of Psychiatry, 1991

23. Goodman R: The extended version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden. Journal of Child Psychology and Psychiatry 40:791–799, 1999

24. Wells MG, Burlingame GM, Lambert MJ: Youth Outcome Questionnaire (Y-OQ). Salt Lake City, Utah, OQ Measures, 1999

25. Singer JD, Willett JB: Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. New York, Oxford University Press, 2003

26. Raudenbush SW, Bryk AS: Hierarchical Linear Models: Applications and Data Analysis Methods. Thousand Oaks, Calif, Sage, 2002

27. Snijders TAB, Bosker RJ: Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling. Thousand Oaks, Calif, Sage, 1999

28. Gibbons RD, Hedeker DR, Davis JM: Estimation of effect size from a series of experiments involving paired comparisons. Journal of Educational Statistics 18:271–279, 1993

29. Serlin RC, Wampold BE, Levin JR: Should providers of treatment be regarded as a random factor? If it ain't broke, don't "fix" it: a comment on Siemer and Joormann (2003). Psychological Methods 8:524–534, 2003

30. Siemer M, Joormann J: Power and measures of effect size in analysis of variance with fixed versus random nested factors. Psychological Methods 8:497–517, 2003

31. Hedeker D, Gibbons RD: Application of random-effects pattern-mixture models for missing data in longitudinal studies. Psychological Methods 2:64–78, 1997

32. Lambert EW, Doucette A, Bickman L: Measuring mental health outcomes with pre-post designs. Journal of Behavioral Health Services and Research 28: 273–286, 2001

33. Nich C, Carroll K: Now you see it, now you don't: a comparison of traditional versus random-effects regression models in the analysis of longitudinal follow-up data from a clinical trial. Journal of Consulting and Clinical Psychology 65: 252–261, 1997

34. Connor Smith JK, Weisz JR: Applying treatment outcome research in clinical practice: techniques for adapting interventions to the real world. Child and Adolescent Mental Health 8:3–10, 2003

35. Sexton TL, Kelley SD: Finding the common core: evidence-based practices, clinically relevant evidence, and core mechanisms of change. Administration and Policy in Mental Health and Mental Health Services Research 37:81–88, 2010

36. Stein BD, Kogan JN, Hutchison SL, et al: Use of outcomes information in child mental health treatment: results from a pilot study. Psychiatric Services 61:1211–1216, 2010