

Clinical Outcome, Consumer Satisfaction, and Ad Hoc Ratings of Improvement in Children's Mental Health

Warren Lambert, Mark S. Salzer, and Leonard Bickman
Vanderbilt University

Mental health clinics and managed care organizations assess treatment effectiveness with consumer satisfaction measures and ad hoc measures of improvement obtained from a single informant; some of these measures are as simple as asking clients whether they improved during treatment. In the present correlational study of 199 treated adolescents, we used a multitrait-multimethod analysis to examine psychometrically measured pathology change (pre- and postassessment of symptoms and functioning), consumer satisfaction, and perceived improvement reported by multiple informants. Confirmatory factor-analytic results indicate that (a) outcome variance due to multiple informants cannot be ignored, (b) consumer satisfaction is unrelated to pathology change, and (c) parent-reported perceived improvement ratings are more akin to satisfaction than to pathology change.

In recent years, concerns about the effectiveness of mental health services have been widely expressed. The cost-containment strategies used by managed care organizations raise consumer fears about a possible decline in the quality of care. Managed care companies (behavioral healthcare organizations; BHOs) wish to show purchasers that their services are effective. Perhaps the only thing that consumers, corporations, and government agree on is the need for mental health agencies and providers to measure the effectiveness of their services.

Mental health professionals can point to meta-analyses of studies as evidence that therapy is beneficial (e.g., Casey & Berman, 1985; Smith, Glass, & Miller, 1980; Weisz, Weiss, Alicke, & Klotz, 1987). However, recent distinctions between lab-based findings (efficacy) and field-based findings (effectiveness) have been sobering. Results of laboratory studies of manualized interventions in ideal treatment conditions may not apply to therapy delivered in the average practice (see Hoagwood, Hibbs, Brent, & Jensen, 1995; Weisz, Donenberg, Han, & Weiss, 1995). Excluding the *Consumer Reports* (CR) survey results ("Mental health," 1995; Seligman, 1995), the effectiveness of psychological therapy in the field has not been demonstrated for either adults or children and adolescents (Weisz, Weiss, & Donenberg, 1992). However, few field-based studies have been conducted (Bickman, 1997).

Purchasers, managed care providers, and consumers need to assess the effectiveness of a particular managed care plan or

provider. As a result, outcome assessment is becoming a common practice for mental health service providers. Just as the clinical researchers before them, evaluators in the field struggle with methodological problems, such as the problem of confounding traits with assessment methods.

The target of most clinical outcome evaluation is pathology change, such as change in symptoms over time. Repeated measurements of symptoms and functioning are used to evaluate the clinical benefit of a particular program or intervention. Clinical effectiveness is determined by (a) the extent to which symptoms decrease and functioning increases, and (b) whether scores for a treated group are significantly better than scores for a comparison group. Unfortunately, repeated measurements are financially expensive and may require more time and effort than clients are willing to provide.

An inexpensive approach to outcome evaluation can be seen in the CR mental health service outcome survey of over 4,000 CR readers ("Mental Health," 1995). Seligman (1995) praised this survey as an ideal study of the effectiveness of psychotherapy. A composite of three items was used to assess clinical effectiveness: (a) satisfaction ("Overall, how satisfied were you with this therapist's treatment of your problems?"); (b) specific improvement ("How much did treatment help with the specific problem that led you to therapy?"); and (c) global improvement (how informants described their "overall emotional state" at the time of the survey compared with the start of treatment). The constructs assessed in the CR survey are satisfaction with treatment (first question) and perceived improvement (second and third questions).

Both satisfaction and improvement have gained popularity in assessments of the effectiveness of mental health services. A recent survey (Bilbrey & Bilbrey, 1995) indicates that over 90% of BHO representatives view customer satisfaction as an important outcome, and of all methods used to evaluate outcome, BHO representatives consider satisfaction to be the most helpful. The collection of satisfaction data has become a high priority for BHOs and mental health agencies and providers (Bilbrey & Bilbrey, 1995; Daniels, Kramer, & Mahesh, 1995;

Warren Lambert, Mark S. Salzer, and Leonard Bickman, Center for Mental Health Policy, Vanderbilt University.

This research was supported by the U.S. Army Health Services Command (DA-DA10-89-C-0013) as a subcontract from the North Carolina Department of Human Resources/Division of Mental Health, Developmental Disabilities and Substance Abuse Services, and Grant R01MH-46136-01 from the National Institute of Mental Health.

Correspondence concerning this article should be addressed to the authors at Center for Mental Health Policy, Vanderbilt University, 1207 18th Avenue South, Nashville, Tennessee 37212. Electronic mail may be sent to lambert@uansv5.vanderbilt.edu.

Trabin, Freeman, & Pallak, 1995). The American Managed Behavioral Healthcare Association (AMBHA) has developed a performance measurement system (PERMS 1.0) that includes numerous indicators related to satisfaction with services (Ross, 1997).

Although satisfaction and service effectiveness are often equated, research examining their relationship has been mixed. One study found a moderate correlation ($r = -.30$) between satisfaction and self-reported pathology change, but no correlation with therapist-rated outcomes (Attkisson & Zwick, 1982). Another study (Deane, 1993) found a moderate correlation between therapist-rated outcomes and satisfaction. Conflicting results were found in two studies measuring symptom change using the Reliable Change Index (Jacobson & Truax, 1991). Ankuta and Abeles (1993) reported greater client satisfaction ratings for those in the clinically significant change group, whereas Pekarik and Wolff (1996) reported no such relationship. Others reported minimal to no relationship between satisfaction ratings and pathology change (Campbell, Ho, Even-son, & Bluebird, 1996; Cauce et al., 1996; Eisen, 1996; H. Mark, personal communication, May, 1996; Ries, Jaffe, Comtois, & Kitchell, 1996) or quality of life (Minsky, 1997).

Typical client satisfaction surveys include items inquiring about the extent to which the consumer perceives that clinical improvement occurred as a result of treatment. We refer to these items as perceived improvement. Perceived improvement involves symptom and functioning areas that are most relevant to each particular client. The measurement of satisfaction and perceived improvement has practical advantages over the measurement of change in pathology. The surveys take little time to complete and are administered only once. The assessment of satisfaction and perceived improvement is considered an inexpensive, yet valid approach to assess overall changes in symptoms and functioning.

A second issue in outcome assessment concerns which data collection methods are used, such as self-report, interview, or behavioral observation, and which informants are asked to assess each outcome trait. Including the informant in an outcome assessment model acknowledges that each respondent has a particular viewpoint; the parent and adolescent surveys are not parallel forms of the same test. If viewpoints were not distinct, asking the parent, "Is your adolescent difficult?" would be the same as asking, "Are you a difficult adolescent?"

The assessment of interrater agreement, especially therapist-client agreement, has a long history in psychotherapy research. One recent literature review (Weiss, Rabinowitz, & Spiro, 1996) showed that the level of agreement between therapist and client ratings of therapy varied greatly across studies. In an extensive review of informant issues in child and family mental health, Northrup (1995) found that the agreement between multiple informants rating child psychopathology is low to moderate (Achenbach, McConaughy, & Howell, 1987; Kazdin, 1994). These interrater discrepancies have led to suggestions that differences in perception among various stakeholders (i.e., clients, therapists) should be studied in their own right rather than disregarded as measurement error (Northrup, 1995; Weiss et al., 1996).

Method variance due to multiple informants may be a source of erroneous results in outcome studies. Large systematic differ-

ences between respondents are treated as error, or shared method variance may inflate positive correlations between measures assessing two seemingly independent traits. Without multiple informants it is unclear whether the relationship is due to shared method variance (i.e., single informant) or traits that are truly correlated. Method variance might account for some moderate correlations between pathology change and client satisfaction.

Campbell and Fiske (1959) clarified the requirements for construct validity by distinguishing between trait variance and method variance. In the present study, traits (constructs of interest) are symptom change and satisfaction with services. Methods (operations used to measure the constructs of interest) include self-ratings by the adolescent client, ratings by a client's parent, and ratings by a trained interviewer. Campbell and Fiske (1959) used the distinction between traits and methods to list a number of requirements for validity. For example, they wrote, "A third common-sense desideratum is that a variable correlate higher with an independent effort to measure the same trait than with measures designed to get at different traits which happen to employ the same method" (p. 83). With this criterion (and others), Campbell and Fiske (1959) strove to protect researchers from misinterpreting method-based correlations as a link between traits. Campbell and Fiske's original article showed how inspecting a multitrait-multimethod (MTMM) matrix of correlations could help determine whether the requirements for convergent and discriminant validity were met, but better methods are available now to replace the inspection of correlation matrices. Confirmatory factor analysis (CFA), also called structural equation modeling (SEM), offers a more objective way to evaluate correlation matrices, to define hypotheses, and to determine which hypotheses offer a better fit to the data. Byrne (1994) and Byrne and Goffin (1993) showed how to program the MTMM in Bentler's (1992) EQS or in Joreskog's LISREL (Byrne, 1989).

The present study uses data from the Fort Bragg Evaluation Project (FBEP), a study of treated children and adolescents in two contrasting systems of mental health care. The FBEP was designed to see whether an experimental continuum of care produced better outcomes and client satisfaction, as well as lower costs, compared with traditional services at two comparison sites. In the FBEP, closely matched groups of clients were treated in the demonstration, a continuum of care, or comparison sites in a quasi-experimental design. Informants included the client, a parent, and a trained interviewer, with additional data from teachers and providers. For the present study, only adolescent clients were used because younger children did not fill out self-reports. The results of the FBEP (Bickman, 1996a; Bickman, 1997; Bickman et al., 1995) and the methods (Bickman, 1996b) of the FBEP appear elsewhere. In the present study, we examine relationships among pathology change (as reported by adolescents, parents, and trained interviewers), client satisfaction (as reported by adolescents and their parents), and perceived improvement (reported by parents). An assumption of this research, one that will be tested empirically, is that the most valid outcome measurement model must include multiple outcome traits (constructs of interest) and multiple methods (multiple informants). We tested three hypotheses: (1) CFA models that distinguish different informants will provide a better fit to the data than those that do not; (2) satisfaction measures

and symptom change measures are distinct; and (3) parent-reported perceived improvement will resemble measures of satisfaction more than measures of pathology change.

Method

Research Participants

The FBEP research sample is described in detail elsewhere (Bickman et al., 1995). The FBEP evaluated the outcome, utilization, and dollar cost of treatment for children ages 5 to 17 who received mental health treatment through the Civilian Health and Medical Program for the Uniformed Services (CHAMPUS). Children were treated in an innovative continuum of care demonstration at Fort Bragg or in traditional reimbursed care at Fort Campbell or Fort Stewart.

The total sample included 498 adolescents from $N = 984$ children of all ages. Cases with any missing parent or child mental health data were excluded from the present study. The overall rate of participation in clinical data collection at Wave 2 was 84% (Bickman et al., 1995, Table 3.13). Although this rate of participation by voluntary subjects is acceptable, a pre-post study using multiple informants makes severe demands on data completeness because a difference score is missing if either the pre- or postmeasure is missing. After dropping cases with incomplete mental health scores, a study sample of 199 adolescents remained. The size of the present sample is adequate by Breckler's (1990) criterion ($N > 100$), and close to the median ($N = 198$ subjects) reported by Breckler's (1990) meta-analysis of published CFA articles. Sample characteristics appear in Table 1.

Categorical data in Table 1 shows a middle class sample with many two-parent homes, a sample different from indigent cases in single-parent homes. Clinical characteristics of the sample appear in Table 2. Presenting problems in this sample most often concern behavior or mood, and the majority have serious emotional disturbance (diagnosis plus impairment).

The average adolescent in the sample had a total problem score in the low-clinical range on the parent-reported Child Behavior Checklist (CBCL; Achenbach, 1991). The sample's mean CBCL total was 64 on

Table 1
Characteristics of the 199 Adolescents Studied

Category	% of 199 adolescents
Male	50.3
Race	
White	64.3
African American	21.1
Two parents in home	84.4
Serious emotional disturbance (diagnosis + impairment)	62.3
Medicated for behavior or emotion	27.9
Parent reports previous child problems	78.1
Parent reports previous child treatment	62.0
Parent reports abuse or violence in home	10.6
Adolescent picked up by police	24.6
Highest parental education included some college	85.4
Site	
Fort Bragg demonstration	52.8
Comparison	47.2
Household income <\$15,000	9.8
Presenting problem conduct or behavior	35.5
Presenting problem anxiety or mood	24.9

Table 2
Clinical Characteristics of the Adolescents Studied

Clinical characteristic	M	SD
Achenbach CBCL total problem <i>t</i> score	64.48	10.08
Achenbach CBCL internalizing <i>t</i> score	61.99	11.37
Achenbach CBCL externalizing <i>t</i> score	64.36	11.07
Achenbach YSR total problem <i>t</i> score	57.51	10.77
Achenbach YSR internalizing <i>t</i> score	55.69	12.26
Achenbach YSR externalizing <i>t</i> score	58.62	10.29
No. of outpatient sessions	14.02	12.90
No. of days in hospital	6.18	16.62
Length of treatment in days	226.77	130.19

Note. CBCL = Child Behavior Checklist. YSR = Youth Self-Report.

a scale in which nonclinical samples have a mean of 50 and standard deviation of 10. Adolescent self-descriptions on the Achenbach's Youth Self-Report (YSR; Achenbach & Edelbrock, 1987) appeared less pathological than the parent descriptions, as did teacher reports, which were missing too often to be used in the present study. The average adolescent in the sample received substantial treatment, with an average of 14 outpatient sessions and 6 hospital days (positively skewed). When service utilization was followed for a year after intake, the average length of treatment extended over 227 days until termination.

Measures

Mental health symptom change and consumer satisfaction were assessed by three kinds of measures: (a) mental health symptom scores, (b) adolescent and parent reports of satisfaction with treatment, and (c) a simple parent-reported perceived improvement scale. A total of 14 variables were used, 7 measuring symptoms, 7 measuring satisfaction. A summary of these measures appears in Table 3.

Symptom change. The symptom scores included assessments of behavior problems, *Diagnostic and Statistical Manual of Mental Disorders*, third edition, revised (*DSM-III-R*; American Psychiatric Association, 1987) psychopathology, and impairment in functioning. In addition to self-reports, trained interviewers rated clients in the FBEP. Interviewers were project employees. Interviews were videotaped, and a sample re-rated; if an interviewer's reliability decreased, they were re-trained.

Included in the assessment of symptoms were (a) self-report checklists by the parent (CBCL; Achenbach, 1991) and by the adolescent client (YSR, Achenbach & Edelbrock, 1987); (b) ratings of symptoms by trained interviewers based on parent report (Parent-Reported Child Assessment Schedule, PCAS; Hodges, McKnew, Cytryn, Stern, & Kline, 1982) or adolescent report (Child Assessment Schedule, CAS; Hodges, Kline, Stern, Cytryn, & McKnew, 1982; Hodges, McKnew et al., 1982); (c) ratings of functioning impairment by trained interviewers (Child and Adolescent Functioning Assessment Schedule, CAFAS, and General Level of Functioning, GLOF; Hodges, 1990; Hodges & Gust, 1995); and (d) a parent-reported rating of caregiver strain resulting from the adolescent's problems (Burden of Care Questionnaire, BCQ; Brannan, Heflinger, & Bickman, 1997). Each of the symptom measures was administered at intake and again at 6 months for all clients, whether or not they were still in treatment.

To measure improvement with the pathology measures, difference scores were used. Difference scores were defined so that for all measures¹ of pathology, a large difference score is good, and a negative

¹ Low GLOF scores indicate poor functioning. However, for consistency, GLOF difference scores were reversed so that a large positive score indicates improvement.

Table 3
Fourteen Measures of Symptoms, Satisfaction, and Perceived Improvement

Type of measure	Name	Content	Occasions	Respondent	Type of validity
Symptom outcome	V1. CBCL total problems	Behavior problems	Pre-post	Parent	Extensive criterion-related validity
	V2. PCAS total pathology	<i>DSM-III-R</i> symptoms	Pre-post	Parent	Recent criterion
	V3. Parental burden	Family difficulties from treated child	Pre-post	Parent	Some criterion
	V4. Youth Self-report	Behavior problems	Pre-post	Adolescent	Extensive criterion
	V5. CAS total pathology	<i>DSM-III-R</i> symptoms	Pre-post	Adolescent	Recent criterion
	V6. CAFAS impairment	Functioning	Pre-post	Interviewer	Some criterion
	V7. GLOF level	Functioning	Pre-post	Interviewer	Extensive criterion
Satisfaction	V8. Global satisfaction	Treated well?	Post only	Parent	Face validity
	V9. Clinic helpful?	Worthwhile to go?	Post only	Parent	Face validity
	V10. Improvement	Child got better?	Post only	Parent	Face validity
	V11. Global satisfaction	Treated well?	Post only	Adolescent	Face validity
	V12. Clinic helpful?	Worthwhile to go?	Post only	Adolescent	Face validity
	V13. Improvement	Child got better?	Post only	Adolescent	Face validity
Perceived improvement	V14. Perceived improvement	How much better?	Post only	Parent	Face validity

Note. CBCL = Child Behavior Checklist; PCAS = Parent-Reported Child Assessment Schedule; CAS = Child Assessment Schedule; CAFAS = Child and Adolescent Functioning Assessment Schedule; GLOF = General Level of Functioning; Pre-post = preassessment to postassessment; *DSM-III-R* = *Diagnostic and Statistical Manual of Mental Disorders*.

difference score means the adolescent's score became worse. Difference scores have been criticized (Cronbach & Furby, 1970) and defended (Newman, 1994); certainly multiwave longitudinal models (Diggle, Laing, & Zeger, 1994) are better than two-wave difference scores for evaluating a clinical trial. However, the present study required a simple operational definition of improvement, and for this purpose, difference scores are adequate.

Measures of client satisfaction. In addition to the symptom remission scores listed in the top half of Table 3, the FBEP evaluated parent and adolescent satisfaction with treatment (Brannan, Sonnichsen, & Heflinger, 1996). These measures appear in the lower half of Table 3 (Satisfaction, V8–V14). The first six measures (V8–V13) were taken from the FBEP satisfaction scales. The last measure (V14, perceived improvement) is a measure of overall outcome similar to that used by CR. The placement of V14 with satisfaction measures reflects our hypothesis that perceived improvement is more similar to satisfaction scales than to symptom scales, an assumption that will be tested later. In the present study, overall measures of satisfaction were based on averages of the service-specific satisfaction scales so that satisfaction scores estimate satisfaction on all services received from intake to 6-month follow-up. Three aspects of satisfaction were used: (a) global satisfaction averaged over all global satisfaction items and all services, (b) "helpfulness" assessed how helpful the informant felt the component of treatment was, and (c) "improved" measured how much the informant felt the adolescent improved during the treatment. Item 3 (improvement items taken from satisfaction scales) will be tested to see if it belongs with satisfaction items, or with mental health symptom changes. The FBEP "perceived improvement" scale was selected for special study because it resembles scales used in the CR study. Its five items simply asked whether the adolescent had improved; the total score was the average of all five items.²

The first seven measures (V1–V7) were difference scores based on measures of pathology. Six of the seven (V1, V2, V4–V7) are established measures backed by extensive research. One measure (V3, parental burden) is a new instrument developed in the FBEP. Burden (V3) and CAFAS Functioning Impairment (V6) are the best predictors of future service utilization in the FBEP.

The satisfaction measures (V8–V13) were taken from the FBEP's satisfaction study (Bickman et al., 1995, p. 121). These items are typical

of those included in client satisfaction measures. The last measure (V14, perceived improvement) is a five-item scale asking the parent whether the adolescent improved; this scale resembles the assessment method used in the CR study. One of the main purposes of the present study is to observe whether perceived improvement behaves more like a measure of symptom change or of satisfaction.

Procedure

In the FBEP, research volunteers were recruited at intake from the population of all children and adolescents treated at the demonstration continuum of care or the comparison's traditional mental health care. Every case received an extensive evaluation including self-report forms and a clinical interview at intake (Wave 1) and 6 months after intake (Wave 2), regardless of whether they were still in treatment. Subjects were the treated adolescent and a parent, often the mother.

Results

Univariate Correlations

Before testing the MTMM hypothesis, simple univariate correlations between measures of symptom change and global satisfaction were examined without factor analysis. First, descriptive statistics for the 14 variables were checked for restriction of range; all variables had ranges in excess of three standard deviations (*SDs*).

The correlations appear in Table 4. For the parent, the correla-

² This scale asked the parent five questions about his or her son or daughter's improvement from intake to 6-month follow up. The first item was "Thinking back to the reasons that brought your child to services, about how much do you think he/she has changed, overall, in the last 6 months? [much better, a little better, stayed the same, a little worse, much worse]." The next four items asked: "... would you say his/her behavior has changed?, ... you think your child would say, ... how ... your child is feeling?, child would say he/she is feeling ...?"

Table 4
Correlations Between Summary Measures of Satisfaction and Symptom Change

Correlation	1	2	3	4	5	6
Parent satisfaction (V8 + V9 + V10)	—					
Parent symptom report (V1 + V2 + V3)	0.38**	—				
Adolescent satisfaction (V11 + V12 + V13)	0.29**	0.14 ^a	—			
Adolescent symptoms (V4 + V5)	0.08	0.19**	0.18**	—		
Interviewer rated functioning (V6 + V7)	0.22**	0.24**	0.18**	0.28**	—	
Parent-reported perceived improvement (V14)	0.65**	0.44**	0.23**	0.13	0.30**	—

^a $p = .06$, nonsignificant.

* $p < .05$. ** $p < .01$.

tion between clinical change on standardized instruments and satisfaction with treatment was 0.38, a moderate correlation consistent with the literature. For the adolescent the correlation was lower but still significant, r (symptom change, satisfaction) = .18. More thorough analyses presented later, however, will suggest that these correlations between satisfaction and symptom change are artifacts of method variance.

The bottom row of Table 4 shows the correlation of the perceived improvement scale with other measures of symptom change and satisfaction. Parent-reported perceived improvement shows significant correlations with many items, including parent-reported satisfaction and symptom change, and adolescent reported satisfaction. In Table 4, V14 (parent perceived improvement), correlates significantly with everything but adolescent reported symptom change. A superficial description of this result might be "a simple perceived improvement scale, correlating with both measures of symptom change and satisfaction, had concurrent validity as a measure of both clinical improvement and satisfaction." The confirmatory factor analysis that follows will examine critically this simplistic interpretation of correlations among measures of satisfaction, clinical change, and perceived improvement to see how well this explanation fits the data.

A matrix of correlations among all 14 variables was calculated using Bentler and Wu's (1993) EQS, a confirmatory factor analysis program. Standard scores with a mean of zero and a standard deviation of one based on the $N = 199$ sample were used to make different units comparable, and all variables were standardized such that a high score was good (much improvement, high satisfaction). Some variables had statistically significant departures from the normal distribution, according to Shapiro and Wilk's W (Shapiro & Wilk, 1965); therefore Satorra-Bentler robust statistics were used, including the robust corrected fit index (CFI*); Byrne, 1994, p 88.). Robust estimates measure the departure from normality and scale parameters and alpha levels accordingly. EQS provides two measures of comparative fit, CFI (normal) and CFI* (robust). EQS made it possible to hypothesize a number of models describing the relationships among clinical difference scores, measures of satisfaction, and perceived improvement and then assess how accurately each model fits the observed correlations. The confirmatory factor analytic models that follow will test whether the superficial conclusions based on Tables 2 and 3 can withstand rigorous examination in a more comprehensive explanatory model.

Byrne (1994, Chapter 6) presented a confirmatory factor

analysis model for multitrait-multimethod problems, illustrated with both the popular LISREL model of Jöreskog (Byrne & Goffin, 1993) and Bentler's (1992) EQS. The present study uses Byrne's EQS version as the starting point for analysis. This analysis makes three assumptions: (a) The 14 measures reflect two constructs ("traits"): mental health symptom change and consumer satisfaction, (b) methods include the three informants: parent, adolescent, and trained interviewer, and (c) a satisfactory model could be used to test two hypotheses: (1) Perceived improvement (V14), and the two improvement items from the satisfaction scales fit better with measures of consumer satisfaction than with measures of mental health symptom change, and (2) Symptom reduction and consumer satisfaction are orthogonal (uncorrelated), not oblique (correlated).

Confirmatory Factor-Analytic Models

Because the hypothesized multitrait-multimethod model is complicated, we tested more parsimonious models first. A complicated model is justified only if it offers a significantly better fit than simpler models. Table 5 shows the results of testing five hypothesized models to account for the correlations among the 14 measures. Once an adequate model is found, specific hypotheses can be tested within it.

Model 1: Outcomes (symptom change and satisfaction) are one. This common factor model states that all 14 measures assess one common factor or construct. This model is used in studies that treat mental health outcome and client satisfaction as "proxies," meaning that if one is expensive you can measure the other instead. In this model, there is only one point of view (usually the parent's). This model is theoretically the most parsimonious model. Unfortunately, Table 5 (row 1) shows that this model has an extremely poor fit with the data. Its robust comparative fit index (CFI*) is only 0.19 on a 0 to 1 scale in which 0.90 is adequate. Model 1 is inaccurate; it must be rejected.

Model 2: Two constructs, symptom change and satisfaction. Model 2 posits a distinction between symptom change and satisfaction but ignores differences among informants. As models in Table 5 add parameters, their fit necessarily improves. In each step an incremental χ^2 was calculated by subtracting the model's total misfit χ^2 from the χ^2 for the preceding step. In Table 5, the misfit χ^2 went down from 463 in row 1 to 412 in row 2. Degrees of freedom decreased from 79 to 78. This leaves an incremental χ^2 difference of 51 with $df = 1$. Because a $\chi^2(1)$

Table 5
Accuracy Estimates From Confirmatory Factor Analysis, Most Parsimonious First

Model	Theory	Normal compar. fit index (CFI)	Robust compar. fit index (CFI*)	Robust $\chi^2(df)$	Incremental significance	Result
Common factor ("It's all one")	All 14 variables are aspects of the same thing. Methods (raters) have negligible differences.	0.46	0.19	463 (79)	(First step)	Extremely poor fit
Two traits, ignore methods	Satisfaction and clinical outcome are distinct. Methods (raters) have negligible differences.	0.53	0.29	412 (78)	$p < .001$	Very poor fit
No traits, just three methods	Raters have distinct viewpoints. Model includes methods only, no traits.	0.76	0.60	331 (77)	$p < .001$	Poor fit
One trait, three methods	Outcome and satisfaction are not distinct. Raters have distinct viewpoints.	0.90	0.86	143 (60)	$p < .001$	Marginal fit
Two traits, three methods (multitrait multimethod)	Outcome and satisfaction are distinct; raters have distinct viewpoints	0.95	0.92	96 (59)	$p < .001$	Good fit

Note. Bentler and Byrne suggest that CFI values less than 0.90 are inadequate. Incremental significance shows whether each step down the table improves the model's fit significantly based on robust chi-square differences from the previous model. Compar. = comparative.

of 3.85 is significant, the incremental χ^2 difference of 51 is not due to chance. Thus, for Hypothesis 2 we conclude that the fit is significantly better than the fit of Hypothesis 1, but, with a robust CFI* of only 0.29, that the fit is still unacceptable.

Model 3: No traits, three methods. In a systematic approach to building a satisfactory model in methodical steps, Hypothesis 3 had to be tested, although it is theoretically unsatisfactory. This hypothesis states that the correlations among the 14 variables can be explained by who reported the data, without considering what they were trying to report. Model 3 fits significantly better than Model 2, showing that method variance makes a contribution, but the model's fit is inadequate. Thus it can be rejected on both empirical and theoretical grounds.

Model 4: One trait, three methods. This model admits that the informants could have different distinct points of view, and it assumes that symptom change and consumer satisfaction are the same. This model offers a much better fit than any before it, with a CFI* = 0.86, but that degree of fit is marginal, because the robust comparative fit index is less than 0.90 (satisfactory).

Model 5: Two traits, three methods. The classical MTMM in Figure 1 has adequate fit (CFI* = 0.92). The MTMM model had a significantly better fit than all the simpler models, as shown in Table 5. Figure 1 shows the MTMM model. The model's structure is a smaller version of Byrne's (1994) MTMM model based on four traits and four methods (self, teacher, parent, and peer). Figure 1 has two traits (symptom change and satisfaction) and three methods (parent, adolescent, and trained interviewer).

Description of the MTMM model. In the initial MTMM model, there are seven measures of symptom change and seven measures of satisfaction. (Treating perceived improvement measures as satisfaction is a hypothesis that will be tested later.) All significant parameters were positive, as expected. The three informants have a modest correlation among them (0.25 to 0.37), suggesting that they are describing the same thing to a limited extent. However, most of what each informant reports is not shared by other informants.

As this study hypothesized, symptom change and satisfaction

were not significantly correlated (normal distribution $r = .06$, $p = .53$, *ns*; robust $r = .10$, $p = .59$). For brevity, only robust estimates will be presented here. The nonsignificant correlation between symptom change and satisfaction suggests that they are

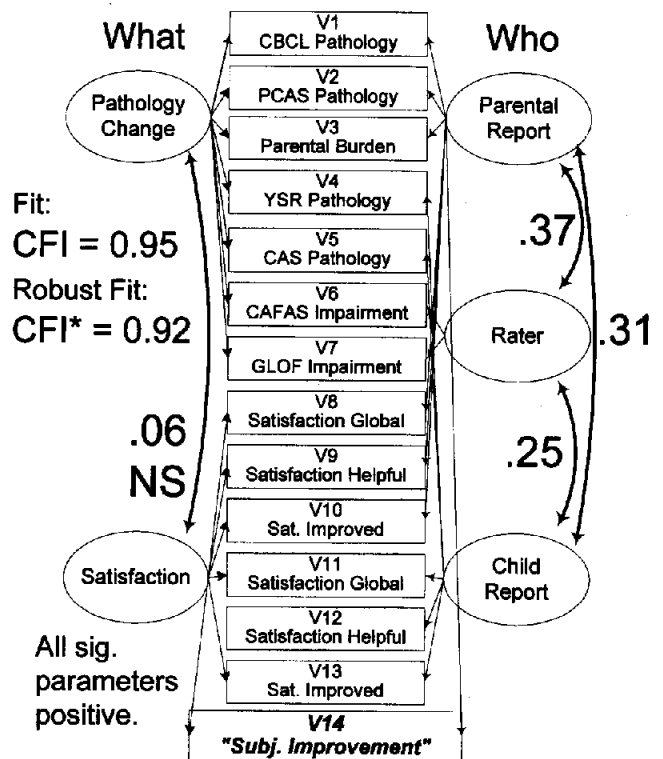


Figure 1. The multitrait multimethod (MTMM) model. CFI = comparative fit index; CBCL = Child Behavior Checklist; PCAS = Parent-Reported Child Assessment Schedule; YSR = Youth Self-Report; CAS = Child Assessment Schedule; CAFAS = Child and Adolescent Functioning Assessment Schedule; GLOF = General Level of Functioning; Sat. = satisfaction; Subj. = subject; Sig. = significant.

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

distinct and unrelated³ and that satisfaction could be equally high regardless of whether the adolescent's symptoms got better or worse. This finding answers the first question of the present study: Are satisfaction and symptom change correlated in this sample? The answer is no.

Role of perceived improvement. In the MTMM model in Figure 1, there were two forms of improvement perceived by the parent. One (V14) was a five-item scale asking the parent whether the adolescent got better or worse over 6 months after intake into treatment. Another (V10) was a similar item from the parent satisfaction scale, with scores averaged over services if there was more than one. In the MTMM model, the five-item perceived improvement scale had a large loading on the parent factor ($\beta = .82, z = 10, p < .01$); however, improvement perceived by the parent was not significantly related to the satisfaction factor ($\beta = .18, z = 1.6, p > .05$). Unlike the five-item perceived improvement scale, the improvement item on the satisfaction scales did load significantly on the satisfaction factor ($\beta = .20, z = 2.0, p < .05$). These loadings are much smaller than those between parent-reported global satisfaction and the satisfaction factor ($\beta = .79, z = 7.74, p < .01$). Although perceived improvement may sometimes be a statistically significant measure of consumer satisfaction, it is not a very good one.

Perceived improvement and outcome. The MTMM model in Figure 1 does not prove that perceived improvement is not a measure of symptom change; perhaps treating improvement both as symptomatic outcome and as satisfaction would improve the model's fit. To test this hypothesis, the MTMM model in Figure 1 was revised. Perceived improvement (V14) was treated as both a symptomatic outcome and as a measure of satisfaction. In Figure 1, an arrow would be drawn connecting V14 to both the pathology change and satisfaction factors. This change did not improve the overall MTMM model's fit, $\chi^2(1, N = 199), = .03 ns$. Furthermore, the loading of parent-perceived improvement on symptom change was only 0.02 ($z = .22, p > .05$); this negligible loading offers no support at all for the hypothesis that the parent's perceived improvement scale is a measure of symptom change.

This test of perceived improvement could be extended to all three measures of perceived improvement, two from satisfaction scales (V10, parent-reported, V13, adolescent-reported) and one from the CR-like perceived improvement scale (V14, parent-reported perceived improvement scale). In Figure 1, arrows could connect V10 (parent-reported improvement), V13 (adolescent-reported improvement), and V14 (parent-reported perceived improvement scale) both to pathology change and satisfaction. Having three more parameters added negligibly to the CFI* (.922 to .926), and the Satorra-Bentler robust χ^2 improvement from these added parameters was nonsignificant, $\chi^2(3, N = 199) = 5.1, ns$. Again, there was no empirical support for the idea that treating perceived improvement as symptom change improved the model.

There was one unexpected finding in the last model. Both parent-reported satisfaction-scale measures of perceived improvement loaded significantly on the satisfaction factor ($\beta_{V10} = .22, z = 2.2, \beta_{V14} = .22, z = 2.1$). In contrast, the adolescent-reported measure of perceived improvement from the satisfaction scales, surprisingly, loaded significantly on symptom

change and nonsignificantly on satisfaction, $r(V13, \text{symptom change}) = .22, z = 2.46$; $r(V13, \text{satisfaction}) = -.04, z = -.5 ns$; these significant loadings were much lower than those for adolescent reported symptom changes (for the YSR, $\beta_{V4} = .56$; for the CAS, $\beta_{V5} = .61$). These results present evidence, admittedly serendipitous, that adolescent-reported perceived improvement may be a measure of symptom change (but not a good one), and that parent-reported improvement may be a measure of satisfaction (but not a good measure of satisfaction either).

A further test of the hypothesis that parent-reported perceived improvement represents satisfaction whereas the same report by an adolescent may represent symptom reduction was tested in another modification of the MTMM model. This modification placed both satisfaction-improvement items (V10, parent; V13 adolescent) in the model in Factor 1, symptom change. For the parent, the resulting loading was nonsignificant, but for the adolescent, it was significant ($z = 2.23, p < .05$). This unanticipated finding, suggesting that adolescent-reported perceived improvement was a measure of symptom change, inspired further post hoc analysis.

Post hoc analysis of adolescent-reported perceived improvement. In response to the preceding results, a new variable (V15) was added to the study's MTMM model, namely the perceived improvement scale as reported by the adolescent (parent version = V14; adolescent version = V15).

When adolescent-reported perceived improvement was added to the MTMM model as a measure of satisfaction, the fit of the MTMM model became significantly worse ($\Delta\chi^2 = 59.3, df = 12, p < .001$). When adolescent-reported perceived improvement was treated as both symptom change and satisfaction in the MTMM model, the model's fit improved ($\Delta\chi^2 = 14.6, df = 1, p < .001$). In this "both" model, the loading of V15 on symptom change was significant, and the loading on satisfaction was nonsignificant. Dropping V15 as satisfaction led to no loss of fit ($\Delta\chi^2 = 0.2, df = 1, p > .05$). In these new 15-variable post-hoc analyses, the correlation between the satisfaction factor and the symptom change factor remained nonsignificant ($r = -.02, p > .05, ns$). It appears that adolescent-reported perceived improvement behaves empirically as a measure of symptom change and not as a measure of satisfaction. This last result may be left as a hypothesis for further research, namely, that self-reported perceived improvement may, for adolescents, be a statistically significant measure of clinical symptom change.

Discussion

The present study had three main findings. First, simplistic models did not fit the data. Adequate fit required distinguishing symptom reduction from client satisfaction, and also recognizing that adolescents, parents, and trained interviewers have separate points of view. These results support the first two hypotheses

³ This lack of correlation is not due to the fact that most children improved during the 6 months after intake. Difference scores had variance, were only moderately skewed. All had ranges greater than five standard deviations. The fact that the average difference score was above zero does not impair the ability of difference scores to correlate with other variables if an empirical connection exists.

stating that multiple traits and multiple methods must be distinguished. Second, although parent-reported perceived improvement has face validity as a measure of symptom change, empirically, the scale behaves more like a measure of client satisfaction. The third result suggested that when satisfaction scales asked adolescents whether they improved, answers were significantly correlated with symptom change. When parents answered the same question, their response correlated with client satisfaction and not with symptomatic change. This last result, found serendipitously, should be tested in further research.

Our data suggest that errors may result from ignoring method (informant) variance in outcome measurement. For example, a statistically inadequate model showed a moderate correlation between parent-reported satisfaction and parent-reported pathology change ($r = .38$); this moderate correlation disappeared ($r = .06$) in the MTMM model. Without the MTMM model, it would have been impossible to determine whether the correlation was due to methods of measurement or due to traits of interest. This result is consistent with the low correlations between informants found in ratings of child pathology (Achenbach et al., 1987; Kazdin, 1994) as well as historic adult studies (e.g., Edwards, Yarvis, Mueller, & Langsley, 1978; Garfield, Prager, & Bergin, 1971) showing low to moderate correlations between therapist and client ratings of success.

The lesson of the MTMM model might be summarized as symmetry, meaning that ideal evaluations should cross each trait with every method in a balanced design. Having multiple informants and multiple measures is feasible in research funded by large grants, but the requirement of symmetry may seem excessive to those evaluating large caseloads in community settings without external funds. A practical solution may be a new breed of assessment tools so inexpensive they can be given to large numbers of cases and informants at a low cost, such as Bickman, Lambert, and Karver's (1997) brief functioning measure. The need to use multiple informants may be affected by the purpose of the measurement. Thus, if the need is to predict service use in program evaluation, fewer informants would be required than in basic research, where a symmetric design may be required to reveal the whole picture.

In the present study, we found no correlation between satisfaction and pathology change, which is similar to findings from many studies (Campbell et al., 1996; Cauce et al., 1996; Eisen, 1996; H. Mark, personal communication, May, 1996; McLellan & Hunkeler, in press; Ries et al., 1996) but not all (e.g., Ankuta & Abeles, 1993; Attkisson & Zwick, 1982). An implication of these results is that satisfaction cannot be used as a proxy for change in psychopathology. This result does not imply that satisfaction scales are without value; the assessment of satisfaction is a reasonable response to demands for a consumer voice in health care evaluations. Unfortunately, current satisfaction scales may be weak in construct validity (Lehow, 1982; Williams, 1994) and may give only the illusion of consumerism (Salzer, 1997; Williams, 1994). More work needs to be done to determine what satisfaction measures assess before satisfaction data would have construct validity.

The conclusions from this study apply to children's mental health services and included three informants (parent, adolescent, and trained interviewer) that are not available in studies involving adults. It would take an adult study using multiple

methods, such as reports by client, spouse, and trained interviewer, to reveal whether the *Consumer Reports* surveys measure satisfaction or pathology change. Because the present study is apparently the first one that attempted to validate the use of satisfaction and perceived improvement, it is possible that perceived improvement behaving as a measure of satisfaction, not psychopathology change, would not be confined to studies of children's mental health.

The *Consumer Reports* survey with face valid ad hoc measures has obvious appeal. To run such a study, the evaluator writes some items, such as "How much did you improve?" On termination, or at a haphazardly chosen time, some clients answer the questions. Unlike other clinical assessments, scores can be tallied without evidence of reliability or validity. Unfortunately, the measurement of clinical effectiveness, with its demands for ratings of multiple traits by multiple informants, may not be easy. A multitrait multimethod longitudinal outcome evaluation may be too expensive for clinics struggling to survive. The present study does not suggest that simple ad hoc measures should never be used, but it does suggest that there is a difference between the results of simple ad hoc perceived improvement-satisfaction surveys and longitudinal studies of clinical outcomes. Users of homespun surveys supported by face validity should acknowledge this limitation in their reports. Those constructing face-valid questionnaires for program evaluation are not relieved of the obligation to present "evidence of the validity and suitability of tests for the purpose of the evaluation and the populations involved" (Standard 12.1, "Program Evaluation," American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1985).

References

- Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Burlington: University of Vermont, Department of Psychiatry.
- Achenbach, T. M., & Edelbrock, C. (1987). *Manual for the Youth Self-Report and Profile*. Burlington: University of Vermont, Department of Psychiatry.
- Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, *101*, 213-232.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.
- Ankuta, G. Y., & Abeles, N. (1993). Client satisfaction, clinical significance, and meaningful change in psychotherapy. *Professional Psychology: Research and Practice*, *24*(1), 70-74.
- Attkisson, C. C., & Zwick, R. (1982). The client satisfaction questionnaire: Psychometric properties and correlations with service utilization and psychotherapy outcome. *Evaluation and Program Planning*, *5*, 233-237.
- Bentler, P. M. (1992). *EQS: Structural equations program manual*. Los Angeles: BMDP Statistical Software, Inc.
- Bentler, P. M., & Wu, E. J. C. (1993). *EQS/Windows user's guide*. Los Angeles: BMDP Statistical Software, Inc.

- Bickman, L. (1996a). A continuum of care: More is not always better. *American Psychologist*, 51(7), 689-701.
- Bickman, L. (1996b). Methodological issues in evaluating mental health services [Special issue]. *Evaluation and Program Planning*, 19(2).
- Bickman, L. (1997). Reactions to Fort Bragg findings: New directions for mental health services research. *American Psychologist*, 52, 562-565.
- Bickman, L., Guthrie, P. R., Foster, E. M., Lambert, E. W., Summerfelt, W. T., Breda, C. S., & Heflinger, C. A. (1995). *Evaluating managed mental health services: The Fort Bragg experiment*. New York: Plenum.
- Bickman, L., Lambert, E. W., & Karver, M. (1997). *A low-cost index of child functioning for predicting service use*. Manuscript submitted for publication.
- Bilbrey, J., & Bilbrey, P. (1995, July/August). Judging, trusting, and utilizing outcomes data: A survey of behavioral health care payers. *Behavioral Health Care Tomorrow*, pp. 62-65.
- Brannan, A. M., Heflinger, C. A., & Bickman, L. (1997). The Caregiver Strain Questionnaire: Measuring the impact of living with a child with serious emotional disturbance. *Journal of Emotional and Behavioral Disorders*.
- Brannan, A. M., Sonnichsen, S., & Heflinger, C. A. (1996). Measuring satisfaction with children's mental health services: Validity and reliability of the satisfaction scales. *Evaluation and Program Planning*, 19(2), 131-141.
- Breckler, S. J. (1990). Applications of covariance structure modeling in psychology: Cause for concern? *Psychological Bulletin*, 107, 260-273.
- Byrne, B. M. (1989). *A Primer of LISREL: Basic applications and programming for confirmatory factor analytic models*. New York: Springer Verlag.
- Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows*. Thousand Oaks, CA: Sage.
- Byrne, B. M., & Goffin, R. D. (1993). Modeling MTMM data from additive and multiplicative covariance structures: An audit of construct validity concordance. *Multivariate Behavioral Research*, 28, 67-96.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multi-trait multi-method matrix. *Psychological Bulletin*, 56, 81-105.
- Campbell, J., Ho, L., Evenson, R. C., & Bluebird, G. (1996, February). *Consumer satisfaction and treatment outcomes*. Paper presented at the Sixth Annual National Conference on State Mental Health Agency Services Research and Program Evaluation, Arlington, VA.
- Casey, R. J., & Berman, J. S. (1985). The outcome of psychotherapy with children. *Psychological Bulletin*, 98(2), 388-400.
- Cauce, A. M., Dyck, D., Hendryx, M., Srebniak, D., Caverly, S., & Setevenson, J. (1996). *Evaluation of SSB6547 Progress Report: Substituting outcomes for process regulations*. Report prepared for the Washington State Division of Mental Health.
- Cronbach, L. J., & Furby, L. (1970). How should we measure "change"--or should we? *Psychological Bulletin*, 74, 68-80.
- Daniels, A., Kramer, T. L., & Mahesh, N. M. (1995, July/August). Quality indicators measured by behavioral group practices. *Behavioral Health Care Tomorrow*, pp. 55-56.
- Deane, F. P. (1993). Client satisfaction with psychotherapy in two outpatient clinics in New Zealand. *Evaluation and Program Planning*, 16, 87-94.
- Diggle, P., Laing, K. Y., & Zeger, S. L. (1994). *Analysis of longitudinal data*. New York: Oxford University Press.
- Edwards, D. W., Yarvis, R. M., Mueller, D. P., & Langsley, D. G. (1978). Does patient satisfaction correlate with success? *Hospital and Community Psychiatry*, 29(3), 188-190.
- Eisen, S. V. (1996, August). Client satisfaction and clinical outcomes: Do we need to measure both? *Behavioral Healthcare Tomorrow*, pp. 71-73.
- Garfield, S. L., Prager, R. A., & Bergin, A. E. (1971). Evaluation of outcome in psychotherapy. *Journal of Consulting and Clinical Psychology*, 37(3), 307-313.
- Hoagwood, K., Hibbs, E., Brent, D., & Jensen, P. (1995). Introduction to the special edition: Efficacy and effectiveness studies of child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology*, 63, 683-687.
- Hodges, K. (1990). *The Child and Adolescent Functional Assessment Scale (CAFAS)*. Unpublished manuscript.
- Hodges, K., & Gust, J. (1995). Measures of impairment for children and adolescents. *Journal of Mental Health Administration*, 22(4), 403-413.
- Hodges, K., Kline, J., Stern, L., Cytryn, L., & McKnew, D. (1982). The development of a child assessment interview for research and clinical use. *Journal of Abnormal Child Psychology*, 10, 173-189.
- Hodges, K., McKnew, D., Cytryn, L., Stern, L., & Kline, J. (1982). The Child Assessment Schedule (CAS) Diagnostic Interview: A report on reliability and validity. *Journal of the American Academy of Child and Adolescent Psychiatry*, 21, 468-473.
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59(1), 12-19.
- Kazdin, A. E. (1994). Psychotherapy for children and adolescents. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change*. New York: Wiley.
- Lebow, J. (1982). Consumer satisfaction with mental health treatment. *Psychological Bulletin*, 91, 244-259.
- Mental health: Does therapy help? (1995, November). *Consumer Reports*, pp. 734-739.
- McLellan, A. T., & Hunkeler, E. (in press). Is patient satisfaction related to patient outcome in alcohol and drug abuse treatment? *Psychiatric Services*.
- Minsky, S. (1997, February). *The relationship between a self-report quality of life instrument and a consumer satisfaction survey*. Paper presented at the Seventh Annual National Conference on State Mental Health Agency Services Research and Program Evaluation, Arlington, VA.
- Newman, F. L. (1994). Criteria for selecting statistical procedures for progress and outcome assessment. In M. E. Maruish (Ed.), *Use of psychological testing for treatment planning and outcome assessment* (pp. 111-134). Hillsdale, NJ: Erlbaum.
- Northrup, D. A. (1995). *The relationship between adolescent and parent reports about family issues*. Unpublished doctoral dissertation, Vanderbilt University, Nashville, TN.
- Pekarik, G., & Wolff, C. B. (1996). Relationship of satisfaction to symptom change, follow-up adjustment, and clinical significance. *Professional Psychology: Research and Practice*, 27(2), 202-208.
- Ries, R. K., Jaffe, C., Comtois, K. A., & Kitchell, M. (1996). *Patient satisfaction and outcome with outpatient integrated dual diagnosis treatment*. Unpublished manuscript.
- Ross, E. C. (1997). Managed behavioral health care premises, accountable systems of care, and AMBHA's PERMS. *Evaluation Review*, 21, 318-321.
- Salzer, M. S. (1997). Consumer empowerment in mental health organizations: Concepts, benefits, and impediments. *Administration and Policy in Mental Health*, 24, 425-434.
- Seligman, M. E. P. (1995). The effectiveness of psychotherapy: The *Consumer Reports* study. *American Psychologist*, 50(12), 965-974.
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52, pp. 115-124.
- Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *Psychotherapy*. Baltimore, MD: Johns Hopkins University Press.

- Trabin, T., Freeman, M. A., & Pallak, M. S. (Eds.). (1995). *Inside outcomes: The national review of behavioral health care outcome programs* (Revised and expanded edition). Tiburon, CA: Centralink Publications.
- Weiss, I., Rabinowitz, J., & Spiro, S. (1996). Agreement between therapists and clients in evaluating therapy and its outcomes: Literature review. *Administration and Policy in Mental Health, 23*, 493–511.
- Weisz, J. R., Donenberg, G. R., Han, S. S., & Weiss, B. (1995). Bridging the gap between laboratory and clinic in child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology, 63*(5), 688–701.
- Weisz, J. R., Weiss, B., Alicke, M. D., & Klotz, M. L. (1987). Effectiveness of psychotherapy with children and adolescents: A meta-analysis for clinicians. *Journal of Consulting and Clinical Psychology, 55*(4), 542–549.
- Weisz, J. R., Weiss, B., & Donenberg, G. R. (1992). The lab versus the clinic: Effects of child and adolescent psychotherapy. *American Psychologist, 47*, 1578–1585.
- Williams, B. (1994). Patient satisfaction: A valid concept? *Social Science and Medicine, 38*(4), 509–516.

Received June 5, 1996

Revision received July 10, 1997

Accepted July 22, 1997 ■

Call for Papers

Psychological Assessment publishes mainly empirical articles concerning clinical assessment. Most articles address the development and validation of psychological assessment instruments. The Editor of *Psychological Assessment* is also interested in publishing articles in the following areas:

- Integrative review of assessment instruments, methods, and topics relevant to clinical assessment;
- The application of basic psychology research to clinical assessment (e.g., cognitive psychology, behavior analysis, neuroscience, social psychology);
- Measurement theory and methods as they apply to clinical assessment;
- Clinical judgment and decision making, including diagnostic assessment and clinical case conceptualization;
- Methods of measuring treatment process and outcome;
- Dimensions of individual differences (e.g., race, ethnicity, age, gender, sexual orientation, economic status) as they relate to clinical assessment;
- Theoretical papers related to clinical assessment; and
- Innovative methods of psychological assessment.

Manuscripts should be submitted to Stephen N. Haynes, Editor, *Psychological Assessment*, Department of Psychology, University of Hawai'i at Manoa, 2430 Campus Road, Honolulu, Hawaii 96822. (See the Instructions to Authors in a copy of the journal, or the Manuscript Submission Information at <http://www.apa.org/journals/pas.html>.)