

# The Client Satisfaction Inventory: Results of an Initial Validation Study

**Steven L. McMurtry**

*University of Wisconsin–Milwaukee*

**Walter W. Hudson**

*Florida State University*

*Objective: To obtain initial results with regard to the reliability and validity of the Client Satisfaction Inventory (CSI), a 25-item scale for measuring general satisfaction with services among clients of human service agencies. Method: The CSI was administered to 329 clients of 11 agencies in six states. Also administered were three other standardized measures and a brief descriptive questionnaire, results from which were used to assess the discriminant validity of the CSI. Results: Findings indicated that both the full version of the CSI and a 9-item short-form version, the CSI-SF, have good to excellent internal consistency. Item analyses also provided some affirmative evidence with regard to the content validity of both versions, and the presence of hypothesized relationships between client satisfaction scores and those of the other instruments offered indications of good discriminant validity for each version. Conclusion: Accountability demands, including pressures associated with managed care, have created a need in many agencies for brief, accurate, and norm-referenced measures of client satisfaction. Although further research is needed, initial results suggest that the CSI and CSI-SF may be useful tools for meeting this need.*

Client satisfaction, as an approach to assessing the quality of services, has a quick and readily apparent appeal. It is easily understood by both clients and providers, can be measured via client self-report (rather than, for example, the more difficult process of behavioral observation), and is a central feature of most definitions of service effectiveness. Indeed, the social validity of human services depends in large measure on client satisfaction. Oddly, however, it has attracted only sporadic attention in literature on evaluating

---

**Authors' Note:** Correspondence should be addressed to Steven L. McMurtry, School of Social Welfare, University of Wisconsin–Milwaukee, P.O. Box 786, Milwaukee, WI 53201; e-mail: mcmurtry@csd.uwm.edu. The Client Satisfaction Inventory and Client Satisfaction Inventory-SF are copyrighted assessment tools and may not be reproduced without permission. To obtain copies or permission to reproduce, please write to WALMYR Publishing Company, P.O. Box 12217, Tallahassee, FL 32317. A free copy of administration and scoring software may be obtained by visiting the following Internet address: <http://www.sypac.com/~walmyr>.



Research on Social Work Practice, Vol. 10 No. 5, September 2000 644-663  
© 2000 Sage Publications, Inc.

social work services. For example, the leading work on practice evaluation by Bloom, Fischer, and Orme (1999) emphasizes client involvement in the evaluation of services but makes no reference to client satisfaction as an outcome measure. Similarly, in their book on practice research, Blythe, Tripodi, and Briar (1994) mention client satisfaction only in the context of assessing the quality of the client-worker relationship.

Still, several of converging trends seem to be directing increased attention to client satisfaction as an indicator of service effectiveness. One of these is the empirical clinical practice (ECP) movement, which advocates that social workers integrate into their services evaluation techniques derived from research methodology. As summarized by Thyer (1996), ECP principles include the use of measurement tools with known validity and reliability, especially the class of brief measures termed *rapid assessment instruments* (RAIs). With respect to client satisfaction, this implies abandoning measurement based solely on practitioner impressions or self-developed surveys in favor of standardized tools with known norms and demonstrated psychometric properties.

The accountability and quality assurance movements have also spurred changes. Beginning with new federal budgeting and monitoring procedures in the late 1960s, concern for accountability in the human services has increasingly required providers to apply objective measures in gauging the impact of their services. In the 1980s and early 1990s, the accountability movement was partly subsumed into the quality assurance movement, which called attention to the need for monitoring organizational processes and outcomes. Young (1995) notes that the development of standardized instruments to measure patient satisfaction in health care settings was one example of innovations that reinforced the spread of quality assurance methods.

Many quality assurance principles were themselves folded into administrative practices touted by supporters of total quality management (TQM). From its roots in the commercial sector, TQM has branched into human services and related fields. It calls for management practices that designate quality as the central criterion guiding all structures and processes in an organization. In most TQM literature, *quality* refers to satisfaction on the part of the consumer. Feigenbaum (1983), an early TQM advocate, defines quality as organizational goods or services that "meet the expectations of customers" (p. 7). Other writers are more specific, arguing that "customer satisfaction is the center or focus of TQM" (Saylor, 1996, p. 2). Moore and Kelly (1996) contend that, in the current era of managed care, documentation of success in satisfying consumers is essential for all human service agencies.

### MEASURING CLIENT SATISFACTION

Whereas studies in various fields have employed client satisfaction as a performance measure, agreement is still lacking as to how it may best be conceptualized. In a review of its use in the mental health field, Lebow (1983) defines client satisfaction as “the extent to which services gratify the client’s wants, wishes, or desires for treatment” (p. 212). In contrast, Young, Nicholson, and Davis (1995) argue that satisfaction with treatment narrows the focus too finely on therapeutic services. They advocate a broader view that addresses satisfaction with multiple types of services, with service integration and coordination, and with the capacity of services to affect the client’s interactions with family members and others. This approach seems better suited to human services that may or may not involve therapy, and it meshes well with a person-in-environment perspective. Young et al. (1995) also note that satisfaction with services may intermingle with satisfaction with the provider, and thus, a broader conceptualization that recognizes this fact is important in measurement development.

Another impediment to theoretical work may be the perception (or misperception) that client satisfaction is so straightforward that measuring it is easy. Larsen, Attkisson, Hargreaves, and Nguyen (1979), for example, note that “a unique feature of client satisfaction research to date is the tendency of investigators to invent their own questionnaires” (p. 198). Despite work by these authors and others to develop standardized measures, recent research still found that “the majority of studies on consumer satisfaction . . . are based on instruments with unknown psychometric properties” (Young et al., 1995, p. 223). McNaughton (1994) also notes this problem in his review of research on satisfaction in children’s special education programs, noting that “analysis of the reported findings was hampered by incomplete descriptions of the measurement tools used. . . . Information on the reliability and validity of the measurement instruments was often unavailable” (p. 26).

Many efforts to create standardized client satisfaction measures have focused on particular problems or populations. Examples of instruments targeted toward specific services are measures of participants’ satisfaction with very brief psychotherapy (Talley, Butcher, & Moorman, 1992), outcomes in HIV counseling (Begley et al., 1994), community mental health services (Love, Caid, & Davis, 1979), inpatient psychiatric services (Holcomb, Adams, Ponder, & Reitz, 1989), private psychiatric hospitals (Glass, 1995), home health care (Westra, Cullen, Brody, & Jump, 1995), rehabilitation counseling (Ju & Thomas, 1987), and nursing home care (Zinn, Lavizzo-Mourey, & Taylor, 1993). Other scales cast satisfaction as a process measure, examining variables such as client-practitioner relationship (Horvath &

Greenberg, 1989) or the quality of individual clinical counseling sessions (Stiles & Snow, 1984).

Still other specialized tools have been designed to gauge parents' satisfaction with services for their children. Examples include scales that obtain parent ratings of community-based services for severely emotionally disturbed children (Rouse, MacCabe, & Toprac, 1994), helping behaviors displayed by mental health professionals when working with children (Cournoyer & Johnson, 1991), services for developmentally disabled children (Clare & Pistrang, 1995), and child and family psychological counseling (Stallard, 1996). A final group of specialized scales allow children and youth to rate their own satisfaction with health care (Rifkin, Wolf, Lewis, & Pantell, 1988; Simonian, Tarnowski, Park, & Bekeny, 1993), psychiatric hospital services (Furst et al., 1993), and general social services (Stüntzer-Gibson, Koren, & DeChillo, 1995).

#### **General Satisfaction Scales**

One instrument that does attempt to measure client satisfaction as a single, broad construct and that offers advantages such as good psychometric properties, brevity, and normative data, is the Client Satisfaction Questionnaire (CSQ-8). Originally developed for use in mental health programs (Larsen et al., 1979), it has since been applied in a variety of other arenas. It is a brief, 8-item scale with very good to excellent internal consistency, based on tested values for coefficient alpha that range from .83 (Roberts & Attkisson, 1983) to .94 (Cox, Brown, Peterson, & Rowe, 1982). It is also unidimensional, consisting of items shown to constitute a single factor in an earlier, 31-item scale (Nguyen, Attkisson, & Stegner, 1983). Findings indicating that it correlates with variables such as program completion provide evidence of its concurrent validity (Attkisson & Zwick, 1982).

The CSQ-8 does have drawbacks, however, such as the fact that it offers only four response options (numbered 1 to 4) for each item, which eliminates the possibility of neutral responses and provides less sensitivity than 5- or 7-point scales. Its response options vary throughout the measure, requiring respondents to adjust to different anchors as they move from item to item. In addition, reverse coding is done by changing the directionality of response anchors rather than by varying item wording from positive to negative, and this, too, may require more careful attention by respondents. Perhaps most important, tests conducted on the scale's readability show it has a grade-level equivalent of 6.8 (Kincaid, Aagard, O'Hara, & Cottrell, 1981). This is acceptable for many purposes but still limits the scale's use with important groups such as younger adolescents or adults with educational deficits.

Researchers involved in the development of the CSQ-8 have published results of other efforts to create a multidimensional satisfaction scale. Drawing on earlier work (Ware & Davies, 1983) that proposed the existence of seven dimensions of patient satisfaction in health care, Greenfield and Attkisson (1989) developed the Service Satisfaction Scale (SSS-30). It was tested in three studies in medical and mental health clinics, and results indicated the presence of not seven but two factors that the authors label *practitioner skill and manner* and *outcome*. Inconclusive results suggested the presence of two other factors termed *office procedures* and *access*. Alphas for the first two subscales across settings and studies averaged .87 and .80, respectively, and .69 and .66 for the second two.

The SSS-30, however, has several features that may greatly limit its usefulness to most social workers. The most important of these is that it was designed to assess "satisfaction with either health or mental health outpatient services" (Attkisson & Greenfield, 1994, p. 402), and it appears appropriate for use only in those settings. Its medical orientation is illustrated by items such as those that ask clients about their satisfaction with "prescription (or non-prescription) of medication" during services and with the "effect of services in helping relieve symptoms" (Greenfield & Attkisson, 1989, p. 273). In addition, tests we conducted on the scale's readability produced a grade-equivalent score of 8.8, and this, too, would limit its use with a substantial number of social work clients.

One other scale measures satisfaction broadly and is targeted specifically toward social work. This is the Reid-Gundlach Social Service Satisfaction Scale (R-GSSSS) (Reid & Gundlach, 1983) which includes 34 items making up three subscales. The subscales measure dimensions of satisfaction termed *relevance* (fit between problem and service provided), *impact* (effect of services on the problem), and *gratification* (effect of services on clients' sense of self-efficacy). In the original study of the R-GSSSS, responses were collected from 166 heads of households served by a Head Start program. Results indicated high alpha (.95) for the full scale. Alphas for the three subscales were .85, .82, and .86, respectively. Our tests show a grade-equivalent reading score of 6.0 for the scale.

Drawbacks of the R-GSSSS include the small number of participants in its original study. It is also lengthier than other measures, and because many items make reference to the social worker, it may not be useful in agencies employing both social work and nonsocial work staff. Of still greater concern is the limited information available on the scale's validity. Neither content nor concurrent validity was addressed in its validation study, and analyses of discriminant and convergent validity were limited to the examination of correlations between subscale scores. In addition, the full scale and/or its

subscales were found to correlate with two of four background variables—ethnicity and marital status. This again raises questions about the scale's discriminant validity because satisfaction theoretically should not correlate with characteristics such as these, and it leaves unresolved the question of whether the scale measures satisfaction or some other construct related to the background variables. The scale appears to have been employed in only one subsequent study (Nickerson, Helms, & Terrell, 1994), which examined 105 undergraduate students. Reliabilities for the three subscales were computed using these data and found to be lower (.73, .69, and .72, respectively) than in the original 1984 study.


This article reports results of a study designed to assess the validity and reliability of a new standardized scale for measuring consumer satisfaction, the Client Satisfaction Inventory (CSI). The CSI was designed to overcome limitations of earlier instruments and to be suitable for use across a variety of clients and services. Findings are reported on psychometric characteristics of a full version of the scale and a short-form version, the CSI-SF.

### The CSI Scale

The full CSI Scale and its short form are paper-and-pencil, self-report measures that most clients should be able to complete in 5 minutes or less. As shown in Figure 1, the full version consists of 25 items, whereas the CSI-SF includes a subset of 9 of these: Items 2, 3, 8, 9, 10, 11, 21, 23, and 24. Items and response options employ simple wording and require only fifth-grade reading skill, as indicated by a score of 5.3 on the Flesch-Kincaid readability index (Kincaid et al., 1981). Clients respond to each item via a 7-point, category-partition scale (Stevens, 1968), in which responses range from 1 (*none of the time*) to 7 (*all of the time*).

Total scores on each scale range from 0 to 100, with higher scores representing higher levels of satisfaction. To lessen the likelihood of response bias, 20 items on the full version are worded in positive terms (in which a response of *all of the time* indicates high satisfaction), and the remaining 5 items are negatively worded. None of the items on the CSI-SF are negatively worded. All items on both versions are arranged randomly.

Scoring the full version requires first that all negatively worded items be reverse scored by changing scores of 1 to 7, 2 to 6, 3 to 5, 5 to 3, 6 to 2, and 7 to 1 (scores of 4 remain unchanged). The numbers of the items that must be scored in reverse are shown beneath the copyright notation at the bottom of the scale (Items 4, 7, 16, 18, and 22 on the full version). After completion of the reverse-scoring process (which is not necessary for the short form), the total score, *S*, is computed as



## CLIENT SATISFACTION INVENTORY (CSI)

This questionnaire is designed to measure the way you feel about the services you have received. It is not a test, so there are no right or wrong answers. Answer each item as carefully and as accurately as you can by placing a number beside each one as follows.

1 = None of the time  
 2 = Very rarely  
 3 = A little of the time  
 4 = Some of the time  
 5 = A good part of the time  
 6 = Most of the time  
 7 = All of the time  
 X = Does not apply

---

1. 6 The services I get here are a big help to me.
2. 6 People here really seem to care about me.
3. 7 I would come back here if I need help again.
4. 1 I feel that no one here really listens to me.
5. 7 People here treat me like a person, not like a number.
6. 6 I have learned a lot here about how to deal with my problems.
7. 2 People here want to do things their way, instead of helping me find my way.
8. 7 I would recommend this place to people I care about.
9. 7 People here really know what they are doing.
10. 5 I get the kind of help here that I really need.
11. 6 People here accept me for who I am.
12. 6 I feel much better now than when I first came here.
13.     I thought no one could help me until I came here.
14. 5 The help I get here is really worth what it costs.
15. 6 People here put my needs ahead of their needs.
16. 1 People here put me down when I disagree with them.
17. 6 The biggest help I get here is learning how to help myself.
18. 2 People here are just trying to get rid of me.
19. 5 People who know me say this place has made a positive change in me.
20. 7 People here have shown me how to get help from other places.
21. 6 People here seem to understand how I feel.
22.     People here are only concerned about getting paid.
23. 6 I feel I can really talk to people here.
24. 6 The help I get here is better than I expected.
25. 7 I look forward to the sessions I have with people here.

---

Copyright © 1994, Steven L. McMurtry
Illegal to Photocopy or Otherwise Reproduce

4, 7, 16, 18, 22.

**Figure 1: The Client Satisfaction Inventory.**

$$S = (\text{Sum } (Y) - N)(100)/[(N)(6)],$$

where  $Y$  is the score for an item and  $N$  is the total number of items correctly completed by the respondent; omitted items and items scored outside the range of 1 to 7 are ignored. An advantage of this system is that it produces

scores ranging from 0 to 100 regardless of the number of items the respondent omits or scores improperly. However, it is recommended that scores based on forms in which the respondent has completed less than 20 items (80%) be discarded (Hudson, 1992).

Figure 1 provides scores from a hypothetical client. Computing a total score for this client would require, first, that Items 4, 7, 16, and 18 be reverse scored as described above. Because the client left Items 13 and 22 blank, they would not be included in the subsequent summing procedure and thus  $N = 23$ . Entering these results into the above formula,  $S = (143 - 23)(100)/[(23)(6)]$ , produces a final rounded score of 87 for the hypothetical client in Figure 1. The same formula is used with the CSI-SF, except that a lower value for  $N$  is used because this version has fewer items.

Because the scale was intended for use in diverse settings, wording in each item that relates to the program or service being evaluated is nonspecific. For example, respondents are asked to rate the "services" or "help" that they were given, but this may apply to many forms of assistance. Similarly, clients are asked to rate the services received "here" or at "this place," which leaves open the option to use the scale in agencies, private practices, or other venues. The verb tense of each item is structured so that it may be used by clients whose services are continuing and by those whose services have ceased (although the time between service cessation and administration of the instrument should not be so great that clients may forget key details affecting their satisfaction). Finally, scale items address cognitive aspects of satisfaction ("I have learned a lot here about how to deal with my problems") and affective aspects ("I feel much better now than when I first came here").

## METHOD

Data were collected via a five-page written form given to clients of human service agencies participating in the study. Four pages consisted of the CSI full version, the Generalized Contentment Scale (GCS), the Index of Self-Esteem (ISE), and the Index of Peer Relations (IPR). The GCS, ISE, and IPR were included as a means of evaluating the content and construct validity of the CSI. Information on their reliability and validity has been documented in prior research (Hudson, 1982). The fifth page of the form was a background questionnaire that requested descriptive information on each respondent. Variables on which data were collected are shown in Table 3.

Data collection began with the identification of agencies willing to administer the instrument to their clients. Some participants were recruited via direct contact by the authors with agencies in the Phoenix metropolitan area



**TABLE 1: Means and Standard Deviations of the CSI and CSI-SF Across Agencies**

Agency Description <sup>a</sup>	State	Client	% of Total	Mean of Total Score	
				CSI	CSI-SF
Crisis counseling	SC	20	6.0	90.1	93.4
Private practice counseling	MI	15	4.5	87.7	91.6
Family/individual counseling	PA	16	4.8	88.4	92.4
Family/individual counseling	AZ	72	21.8	87.0	88.9
Child/family counseling	AZ	16	4.8	87.5	92.1
Adolescent residential treatment	AZ	7	2.1	67.5	67.6
Parent education	AZ	14	4.2	81.8	85.7
Family/individual counseling	TX	38	11.5	84.5	87.8
Older adult programs	WI	102	30.8	83.5	86.0
Family/individual counseling	AZ	21	6.3	89.4	92.9
Veterans' counseling	AZ	10	3.0	81.3	84.5
Overall scale means			85.3	88.1	
Scale standard deviations			12.0	12.4	
<i>F</i> ratio			3.16*	3.51*	
$\eta^2$			.09	.10	

NOTE: CSI = Client Satisfaction Inventory, CSI-SI = Client Satisfaction Inventory-Short Form.

\* $p < .001$ .

and elsewhere that provided field placements to social work students. Others were recruited through a notice sent to members of the social work measurement group, a mutual-interest forum on the Internet. A total of 11 agencies were identified by these means, and their locations and primary services are shown in Table 2. Participating agencies were given a copy of the data-collection instrument and a written statement detailing the purpose of the study. They were asked to administer the instrument to clients for whom such assessments were deemed appropriate, in tandem with any other measurement procedure already being used. Completion of any part of the data collection instrument was voluntary on the part of each client.

An assumption inherent in the use of either the CSI or its short form is that clients to whom it is given have received services for a sufficient time to develop informed opinions of their satisfaction. Because services vary widely in type and duration, however, it is not feasible to set a minimum service period prior to which each version should not be given. For this study, participating agencies were asked to administer the data collection package to clients who were at the end of services, close to the end of their services, or were ongoing clients who met the criterion of being able to form a knowledgeable opinion of their services thus far. As instruments were completed,

**TABLE 2: Client Background Variables and Correlations With Total Scores of the CSI and CSI-SF**

Variable	Coding	M <sup>a</sup>	SD	Correlation With Total Score	
				CSI	CSI-SF
Age		46.43	20.40	-.05	-.04
Gender	Male = 0 Female = 1	.77	.42	.23	.17
Race/ethnicity	White = 0 Non-White = 1	.14	.35	-.00	-.04
Years of education		13.01	2.94	.12	.10
Household income per month <sup>b</sup>		1.83	1.56	-.07	-.05
Marital status	Unmarried = 0 Married = 1	.39	.49	.01	.00
Number of times married		1.07	0.85	.11	.12
Years with current spouse		16.73	17.05	-.13	-.14
Number of children		2.16	2.07	.11	.10
Persons in household		2.88	1.69	-.04	-.03
Reason for seeking services	Involuntary = 0 Voluntary = 1	.87	.34	.13	.13
Mean correlations				.04	.03

NOTE: CSI = Client Satisfaction Inventory, CSI-SF = Client Satisfaction Inventory-Short Form.

a. For dichotomous variables (those coded 0 or 1), the value for the mean can be multiplied by 100 and interpreted as the percentage of cases in the group coded 1. For example, 77% of clients in the study were female (.77 × 100).

b. House income per month is measured in thousands of dollars.

they were sent to the authors for scoring, identified only by agency-assigned code numbers. Results were then returned to each agency, and a copy of the data was retained by the authors. When data collection ceased, a total of 329 completed instruments had been received from the 11 participating agencies.

## RESULTS

### Reliability

Coefficient alpha is a measure of the internal consistency of a scale that has a number of desirable qualities as an estimate of reliability (Nunnally & Bernstein, 1994). The value of alpha for the full CSI scale was found to be .93. Data on item-total correlations were analyzed to identify a subset of

**TABLE 3: Item-Total Correlations Between CSI Items and Total Scores for the CSI, CSI-SF, GCS, ISE, and IPR**

<i>CSI Item</i>	<i>CSI<sup>a</sup></i>	<i>CSI-SF<sup>a</sup></i>	<i>GCS</i>	<i>ISE</i>	<i>IPR</i>
1	.62		-.20	-.15	-.16
2	.57	.55	-.07	-.04	-.09
3	.71	.71	-.03	-.03	-.07
4	.46		-.01	.02	-.04
5	.42		-.12	-.10	-.16
6	.62		-.10	-.06	-.10
7	.44		-.03	-.05	-.09
8	.66	.67	-.14	-.14	-.13
9	.57	.57	-.13	-.12	-.12
10	.74	.74	-.12	-.12	-.12
11	.56	.54	-.11	-.08	-.21
12	.50		-.31	-.18	-.15
13	.38		-.03	-.02	-.20
14	.61		-.23	-.19	-.19
15	.49		-.12	-.16	-.16
16	.50		.15	-.03	-.03
17	.57		-.07	-.07	-.07
18	.51		-.03	.01	-.01
19	.51		-.14	-.14	-.11
20	.50		-.01	-.03	-.08
21	.70	.69	-.09	-.11	-.12
22	.51		-.18	-.15	-.22
23	.67	.69	-.14	-.09	-.11
24	.64	.57	-.19	-.13	-.16
25	.69		-.12	-.06	-.10
Mean	.57	.64	-.10	-.09	-.12

NOTE: CSI = Client Satisfaction Inventory, CSI-SF = Client Satisfaction Inventory-Short Form, GCS = Generalized Contentment Scale, ISE = Index of Self-Esteem, IPR = Index of Peer Relations.

a. Adjusted for item-self correlations.

items to be included the CSI-SF. After the final nine such items had been selected, the value of alpha derived for the short form was .89.

It should be noted that alpha is based on interitem correlations within a scale. These values can vary as a function of differences in sample or population standard deviations. A statistic that is not influenced by such differences, the standard error of measurement (SEM), was also computed for both the CSI and CSI-SF. The results showed values of 3.16 for the full version and 4.11 for the CSI-SF.

On average, a client's obtained score will fall within plus or minus two SEMs of his or her hypothetical true score about 95% of the time, hence good

**TABLE 4: Correlations of Total Scores Between the CSI and CSI-SF, and Their Correlations With Total Scores of the GCS, ISE, and IPR**

<i>Total Score of Comparison Scale</i>	<i>Correlation With Total Score</i>	
	<i>CSI</i>	<i>CSI-SF</i>
CSI-SF	.94	
GCS	-.19	-.16
ISE	-.16	-.13
IPR	-.20	-.17

NOTE: CSI = Client Satisfaction Inventory, CSI-SF = Client Satisfaction Inventory-Short Form, GCS = Generalized Contentment Scale, ISE = Index of Self-Esteem, IPR = Index of Peer Relations.

measures tend to have a large alpha coefficient and a small SEM. A useful rule of thumb concerning the SEM is that its value should equal no more than 5% of the range of possible scores (Hudson, 1990). For 100-point scales such as these, this would mean that the SEM should have a value less than 5, which was found to be true for both CSI scales. With regard to alpha, Rosenthal (1994) notes that scales with values of .80 or higher are appropriate for use in virtually all circumstances, except when clinical inferences are to be made with regard to a single individual, in which case alphas of .90 or higher are desirable. Based on their obtained values for alpha and SEM, and given that they are not individual assessment tools but are designed to measure satisfaction in aggregate, the CSI and its short form appear to meet necessary criteria with regard to their measurement error characteristics.

#### Differences Across Agencies

Before discussing further results, it is important to address the characteristics of the agencies that provided the data. Table 1 briefly describes each agency in the study (or participating program in multiprogram agencies). Shown in the column next to each agency is the two-letter abbreviation for the state where it is located. As the table indicates, counseling-related programs were most common, with 8 of the 11 programs providing that type of service. The residential treatment center may also be appropriate for inclusion in this category. With respect to clients, a more even division occurred between those in counseling programs and those in other types, with the older adult program and parent education program together accounting for 35% of participating clients.

As will be discussed, it was hypothesized that total scores on the CSI and CSI-SF would *not* be associated with variations across clients in background

variables such as age or other characteristics. However, it was hypothesized that CSI and CSI–SF total scores *would* be significantly different across programs or agencies because variation in satisfaction with services received from these programs is what the scales were designed to measure. The latter hypothesis was tested via computation of an omnibus  $F$  test, for which results are shown at the bottom of Table 1. Findings support the hypothesis that agencies varied from one to the next in how they were rated by clients on both versions, as indicated by the statistically significant  $F$  values for both the CSI and CSI–SF. The  $\eta^2$  statistic at the bottom of each column indicates that from 9% to 10% of variation in the total score of each version is accounted for variation across agencies (i.e., by multiplying each  $\eta^2$  value by 100).

Although agencies overall were expected to differ in their clients' level of satisfaction, it was important that the CSI and CSI–SF not produce results that favored one type of agency over another because this might indicate that the scales measured not satisfaction but some other construct having to do with the type of service provided. To test this, it was hypothesized that the mean score on both CSI versions would not differ significantly when comparing agencies that provide counseling services with those offering other types. Results of  $t$  tests on differences between means of the two groups indicated that no such differences were present for either version, thus this hypothesis was supported.

With respect to norms, the mean for the full CSI Scale across clients in all agencies was 85.3, and its standard deviation was 12.0 (as shown at the bottom of Table 1). The mean for the CSI–SF was slightly higher than the full scale, at 88.1, whereas its standard deviation of 12.4 was very similar.

### Client Variables

With regard to characteristics of individual clients participating in the study, Table 2 summarizes descriptive results for the 11 background variables on which information was collected. As can be seen, respondents were on average middle-aged, although they ranged in age from 12 to 89 years. They were also mostly female (77%) and had completed a mean of 13.0 years of school, ranging from 6 to 22 years. Their mean monthly take-home household income of \$1,836 supported an average of 2.9 persons. Clients who identified themselves as White accounted for 86% of the sample, Hispanic and African American clients accounted for about 5% each, and the remaining 4% were from other or unreported racial and ethnic categories.

Information was also collected on respondents' marital status, and results indicated that 39% were currently married. Of those not married, 44% were divorced or separated, 43% were single and had not been married previously,

and the remainder were widowed or of unreported marital status. On average, the respondents had been married once (ranging from never to five times), and those who were currently married had been with their spouse an average of slightly less than 17 years. Seventy-four percent of respondents reported having at least 1 child, with a median of 2 and a maximum of 11. Finally, most respondents were voluntary clients, with 87% reporting receiving services from the agency they were evaluating because they wanted to rather than because they were required to.

### **Content Validity**

Content validity relates to the question of whether each item in an instrument was sampled from the domain of the construct in question—client satisfaction in this case. Some authors have proposed the existence of separate dimensions within the domain of satisfaction, but research has not clearly supported this. As noted earlier, for example, Reid and Gundlach (1983) found evidence of three hypothesized dimensions of satisfaction, but their study was based on a small number of participants and provided no supplementary findings on content or concurrent validity. In a much larger study of the SSS-30, Greenfield and Attkisson (1989) attempted to measure seven dimensions of satisfaction but found good evidence of only two. Their unidimensional CSQ-8, meanwhile, appears to be a more versatile and broadly applicable measure than the multidimensional SSS-30.

Satisfaction was thus approached in our study as a single, global construct. To test the adequacy with which the domain of this construct was sampled by the CSI and CSI-SF, item analyses were conducted to determine whether items on each version contribute significantly to its total score. This process involved correlating each scale item with the total score for that scale and with the total scores for the GCS, ISE, and IPR. Results of this process are shown in Table 3. When correlating any item with an overall scale score, it is important to exclude that item from the computation of the total score because item-self correlations will tend to exaggerate item-total correlations and thus the overall estimate of content validity. The results shown in the second and third columns of Table 3 are corrected for item-self correlations on the CSI and CSI-SF.

As the table shows, the corrected item-total correlations for both CSI scales were moderate to large, with a mean item-total correlation of .57 for the full scale and .64 for the short form. Each correlation was also statistically significant at the .01 level. Finally, as expected, the mean correlation between each CSI item and the total score for each of the other three scales was negligible (GCS =  $-.10$ , ISE =  $-.09$ , and IPR =  $-.12$ ). These results suggest that

items on the CSI were sampled from the same, single domain. In addition, CSI items appear unlikely to have come from the domain of any of the constructs addressed by the other three scales. These results alone are insufficient to support definitive conclusions, but they conform to steps recommended by Nunnally and Bernstein (1994) for providing initial affirmative evidence with regard to the content validity of an instrument.

### **Construct Validity**

Construct validity, as defined by Campbell and Fiske (1959), consists of both convergent and discriminant validity. These two aspects of validity involve the assumption that a valid measure will correlate poorly with variables believed to be unrelated to the construct (discriminant validity) but moderately to strongly with related variables (convergent validity). Results from Table 3 already discussed with regard to content validity are also relevant to construct validity. At the item level of analysis, the total scale score is one variable with which each item should be expected to correlate strongly, hence the values of .57 and .64 for the average correlation between each CSI and CSI-SF item, and the total scale scores is consistent with what would be expected for measures with good convergent validity. In addition, the item-total correlations shown in the second and third columns of Table 3 are identical to factor loadings one would obtain when employing the multiple-group method of factor analysis, a confirmatory technique. The high values for these factor loadings are again useful but not final evidence of the convergent validity of the scales. In addition, each item correlated poorly with variables with which they should not correlate, as shown by the low correlations of CSI items with total scores on the GCS, ISE, and IPR scales. This provides some evidence of discriminant validity at the item level of analysis.

At the subscale level of analysis, discriminant validity can be addressed by determining whether the total score for each CSI version correlates with client background variables. Research on the issue of whether clients' satisfaction is related to demographic characteristics has produced mixed results. Talley et al. (1992) note that no evidence has been found of a relationship between client satisfaction and age or socioeconomic indicators such as education or income. Race and ethnicity is less clear, however, with some studies (Larsen et al., 1979; Reid & Gundlach, 1983) showing non-White clients to be less satisfied than White clients, whereas another study (Stüntzer-Gibson et al., 1995) found no effect for race or ethnicity. With regard to marital status, Reid and Gundlach (1983) found that married or widowed clients were more satisfied than others, but results reported by Larsen et al. (1979) indicated no relationship between marital status and satisfaction. In a comprehensive

review, Tanner (1981) notes that some studies find females to be more satisfied than males, but most have detected no relationship between satisfaction and gender. Larsen et al. found that income had no apparent effect on satisfaction. Finally, contrary to what might be expected, research indicates that voluntary and involuntary clients do not differ in their ratings of satisfaction with services (Spensley, Edwards, & White, 1980). No other background variable tested in this study appears to have been addressed in prior research.

Because none of the demographic characteristics examined in the above studies was consistently found to be associated with client satisfaction, it was hypothesized that CSI and CSI-SF scores would not correlate with any of the background variables addressed here. It was further reasoned that an absence of meaningful correlations between the two versions' total scores and these variables would be evidence of each scales' discriminant validity at the subscale level. The correlation between the total score of the CSI and CSI-SF and each background variable, along with the mean correlation of each scale with all background variables, is shown in Table 2. As can be seen, the average correlation is only .04 between each total scale score and each background variable, and gender was the sole variable found to correlate even weakly with either of the scale scores. These results thus appear to provide supportive evidence with regard to the discriminant validity of the CSI and CSI-SF at the subscale level of analysis.

Further evidence of discriminant validity is shown in the pattern of correlations between the total scores of the two versions and total scores of the GCS, ISE, and IPR. These correlations are shown in Table 4. It was hypothesized that no correlation would exist between total scores of the two CSI scales and those of the GCS, ISE, and IPR. The reason for this expectation is that an accurate measure of satisfaction should not be affected by, for example, the extent to which respondents are feeling somewhat depressed or suffering a poor view of themselves that day. The low values of the correlation coefficients (ranging from  $-.13$  to  $-.20$ ) support this hypothesis.

## DISCUSSION

Certain limitations of this study should be noted before considering what conclusions may be drawn from its results. First, the aim of the study was to include as great a diversity as possible of both agencies and clients. This was done to increase the likelihood of obtaining a heterogeneous sample in which estimates of the reliability and validity of the CSI would not be artificially attenuated by small standard deviations and interitem correlations within the scale. Still, it is important to note that the 11 agencies participating in the



study constituted only a convenience sample of human service organizations, thus their representativeness of the population of all such agencies is unknown. Also unknown is the extent to which clients on whom each agency collected information are representative of the population of clients in that agency. Because of this, sampling error with regard to the mean and standard deviation of the 329 clients included in this study cannot be determined, and the information provided on client norms must be viewed as only an initial estimate.

A more general limitation concerns the use of consumer satisfaction as an outcome measure. Client satisfaction is a self-report measure, and as Martin and Kettner (1996) note, "Client self-reporting by its very nature is not precise. Asking clients to self-report on their quality-of-life changes incurred as a result of having participated in human service programs is perhaps even less precise" (p. 98). In addition, this type of measure may be more reactive and more susceptible to manipulation or distortion than other types (Hersen & Barlow, 1976). For example, a client who benefited from an agency's services but disliked a particular employee or other aspect of the organization might report dissatisfaction, yet this would not accurately reflect the outcome of services. Conversely, clients who benefit little but wish to curry favor might give responses that exaggerate service effectiveness. Client satisfaction measurement should thus be seen as one part of a comprehensive outcome-assessment strategy that includes measures of client change on target problems as well as posttest-only or retrospective indicators of their satisfaction.

Still, client satisfaction remains a core element in most approaches to assessing agency effectiveness, and a major problem with the way satisfaction has been measured in human services in the past has been the tendency to use hastily developed and untested measures that yield confusing or unconvincing results. The CSI and CSI-SF, in contrast, are quick, straightforward, and usable with clients having limited reading skills. Results indicate that they have very good to excellent internal consistency and a low SEM. Initial evidence is also supportive with regard to their content and construct validity, although further research is needed in this area. The total scores of each version do not correlate with any of the background variables measured, which means that they can be used with clients who differ in age, gender, ethnicity, and other characteristics. Finally, scores on the CSI and its short form were found to be unrelated to agency or service type, thus they appear appropriate for use across a range of clients and organizational settings. The CSI-SF in particular is also brief enough to be included as part of a client satisfaction survey in which it is combined with self-developed items addressing unique features of that agency.

The CSI scales also appear to have certain advantages over existing standardized measures such as the CSQ-8 and the R-GSSSS. They use a familiar 100-point scale scoring that lends itself to easy interpretation. As compared to the CSQ-8, they employ standard response anchors that can provide increased understandability and allow them to be incorporated into larger surveys using similarly scaled items. Initial evidence suggests that they are also similar to the CSQ-8 in having strong psychometric properties, but they may be used with clients whose reading skills are almost two grade levels lower. As compared to the R-GSSSS, the CSI scales are briefer, have more evidence of validity, can be applied to both social work and other services, and were tested on a larger and more diverse sample from which to derive norms. Additional research on the concurrent validity of the CSI scales is needed in which scores are correlated with other presumably valid measures and/or with variables such as program continuation versus dropout. Further research is also necessary to determine whether the results reported here are replicable in other settings, but initial evidence suggests that the CSI and its short form are potentially useful new measures for assessing client satisfaction.

## REFERENCES

- Attkisson, C. C., & Greenfield, T. K. (1994). Client Satisfaction Questionnaire-8 and Service Satisfaction Scale-30. In M. E. Mariush (Ed.), *The use of psychological testing for treatment planning and outcome assessment* (pp. 402-420). Hillsdale, NJ: Lawrence Erlbaum.
- Attkisson, C. C., & Zwick, R. (1982). The Client Satisfaction Questionnaire: Psychometric properties and correlations with service utilization and psychotherapy outcome. *Evaluation and Program Planning, 6*, 299-314.
- Begley, K., Ross, M. W., Austin, P., Casey, K., Collins, P., Hennings, G., Agriesti, L., & Marshall, K. (1994). Development and evaluation of an inventory for rating client satisfaction with outcome in HIV counseling: The Albion Center Scale. *Patient Education and Counseling, 24*, 341-345.
- Bloom, M., Fischer, J., & Orme, J. G. (1999). *Evaluating practice: Guidelines for the accountable professional* (3rd ed.). Boston: Allyn & Bacon.
- Blythe, B., Tripodi, T., & Briar, S. (1994). *Direct practice research in human service agencies*. New York: Columbia University Press.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.
- Clare, L., & Pistrang, N. (1995). Parents' perceptions of portage: Towards a standardized measure of parent satisfaction. *British Journal of Learning Disabilities, 23*, 110-117.
- Cournoyer, D. E., & Johnson, H. C. (1991). Measuring parents' perceptions of mental health professionals. *Research on Social Work Practice, 1*, 399-415.
- Cox, J. B., Brown, T. R., Peterson, P. D., & Rowe, M. M. (1982). A report on a statewide community mental health center outcome study. *Community Mental Health Journal, 18*, 135-150.

- Feigenbaum, A. (1983). *Total quality control* (3rd ed.). New York: McGraw-Hill.
- Furst, D., Boever, W., Cohen, J., Dowd, T., Daly, D., & Criste, T. (1993). Implementation of the Boys Town psychoeducational treatment model in a children's psychiatric hospital. *Hospital and Community Psychiatry, 44*, 863-868.
- Glass, A. P. (1995). Identifying issues important to patients on a hospital satisfaction questionnaire. *Psychiatric Services, 46*, 83-85.
- Greenfield, T. K., & Attkisson, C. C. (1989). Progress toward a multifactorial Service Satisfaction Scale for evaluating primary care and mental health services. *Evaluation and Program Planning, 12*, 271-278.
- Hersen, M., & Barlow, D. H. (1976). *Single case experimental designs: Strategies for studying behavior change*. New York: Pergamon.
- Holcomb, W. R., Adams, N. A., Ponder, H. M., & Reitz, R. (1989). The development and construction validation of a consumer satisfaction questionnaire for psychiatric inpatients. *Evaluation and Program Planning, 12*, 189-194.
- Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology, 36*, 223-233.
- Hudson, W. W. (1982). *The clinical measurement package: A field manual*. Chicago: Dorsey.
- Hudson, W. W. (1990). *The MPSI technical manual*. Tempe, AZ: WALMYR.
- Hudson, W. W. (1992). *The WALMYR assessment scales scoring manual*. Tempe, AZ: WALMYR.
- Ju, J. J., & Thomas, K. R. (1987). The accuracy of counselor perceptions of client work values and client satisfaction. *Rehabilitation Counseling Bulletin, 30*, 157-166.
- Kincaid, J. P., Aagard, J. A., O'Hara, J. W., & Cottrell, L. W. (1981). Computer readability editing system. *IEEE Transactions on Professional Communications, 24*, 38-41.
- Larsen, D. L., Attkisson, C. C., Hargreaves, W. A., & Nguyen, T. D. (1979). Assessment of client/patient satisfaction: Development of a general scale. *Evaluation and Program Planning, 2*, 197-207.
- Lebow, J. L. (1983). Research assessing consumer satisfaction with mental health treatment: A review of findings. *Evaluation and Program Planning, 6*, 211-236.
- Love, R. E., Caid, C. D., & Davis, A., Jr. (1979). The user satisfaction survey: Consumer evaluation of an inner city community mental health center. *Evaluation and the Health Professions, 2*, 42-54.
- Martin, L. L., & Kettner, P. M. (1996). *Measuring the performance of human service programs*. Thousand Oaks, CA: Sage.
- McNaughton, D. (1994). Measuring parent satisfaction with early childhood intervention programs: Current practices, problems, and future perspectives. *Topics in Early Childhood Special Education, 14*, 26-48.
- Moore, S. T., & Kelly, M. J. (1996). Quality now: Moving human services organizations toward a consumer orientation to service quality. *Social Work, 41*(1), 33-40.
- Nguyen, T. D., Attkisson, C. C., & Stegner, B. L. (1983). Assessment of patient satisfaction: Development and refinement of a Service Evaluation Questionnaire. *Evaluation and Program Planning, 6*, 299-314.
- Nickerson, K. J., Helms, J. E., & Terrell, F. (1994). Cultural mistrust, opinions about mental illness, and black students' attitudes toward seeking psychological help from white counselors. *Journal of Counseling Psychology, 41*, 378-385.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York: McGraw-Hill.
- Reid, P. N., & Gundlach, J. H. (1983). A scale for the measurement of client satisfaction in the social services. *Journal of Social Service Research, 7*, 37-54.

- Rifkin, L., Wolf, M., Lewis, C., & Pantell, R. (1988). Children's perceptions of physicians and medical care: Two measures. *Journal of Pediatric Psychology, 13*, 247-254.
- Roberts, R. E., & Attkisson, C. C. (1983). Assessing client satisfaction among Hispanics. *Evaluation and Program Planning, 6*, 401-413.
- Rosenthal, J. A. (1994). Reliability and social work research. *Social Work Research, 18*, 115-121.
- Rouse, L. W., MacCabe, N., & Toprac, M. G. (1994, March). *Measuring satisfaction with community-based services for severely emotionally disturbed children: A comparison of questionnaires for children and parents*. Paper presented at the Seventh Annual Research Conference for a "System of Care" for Children's Mental Health, Tampa, FL.
- Saylor, J. H. (1996). *TQM simplified: A practical guide* (2nd ed.). New York: McGraw-Hill.
- Simonian, S., Tarnowski, K., Park, A., & Bekeny, P. (1993). Child, parent and physician perceived satisfaction with pediatric outpatient visits. *Developmental and Behavioral Pediatrics, 14*, 8-12.
- Spensley, J., Edwards, D., & White, E. (1980). Patient satisfaction and involuntary treatment. *American Journal of Orthopsychiatry, 50*, 725-727.
- Stallard, P. (1996). Validity and reliability of the Parent Satisfaction Questionnaire. *British Journal of Clinical Psychology, 35*, 311-318.
- Stevens, S. S. (1968). Ratio scales of opinion. In D. K. Whitla (Ed.), *Handbook of measurement and assessment in behavioral science* (pp. 171-199). Reading, MA: Addison-Wesley.
- Stiles, W. B., & Snow, J. S. (1984). Counseling session impact as seen by novice counselors and their clients. *Journal of Counseling Psychology, 31*, 3-12.
- Stüntzer-Gibson, D., Koren, P. E., & DeChillo, N. (1995). The Youth Satisfaction Questionnaire: What kids think of services. *Families in Society, 76*, 616-624.
- Talley, J. E., Butcher, A. T., & Moorman, J. C. (1992). Client satisfaction with very brief psychotherapy. In J. L. Talley (Ed.), *The predictors of successful very brief psychotherapy* (pp. 46-84). Springfield, IL: Charles C. Thomas.
- Tanner, B. A. (1981). Factors influencing client satisfaction with mental health services: A review of quantitative research. *Evaluation and Program Planning, 4*, 279-286.
- Thyer, B. A. (1996). Guidelines for applying the empirical clinical practice model to social work. *Journal of Applied Social Sciences, 20*, 121-127.
- Ware, J. E., & Davies, A. R. (1983). Behavioral consequences of consumer dissatisfaction with medical care. *Evaluation and Program Planning, 6*, 291-298.
- Westra, B. L., Cullen, L., Brody, D., & Jump, P. (1995). Development of the Home Care Client Satisfaction Instrument. *Public Health Nursing, 12*, 393-399.
- Young, A. T. (1995). Quality assurance. In *Encyclopedia of social work* (19th ed., Vol. 3, pp. 2016-2018). Washington, DC: National Association of Social Workers.
- Young, S. C., Nicholson, J., & Davis, M. (1995). An overview of issues in research on consumer satisfaction with child and adolescent mental health services. *Journal of Child and Family Studies, 4*, 219-238.
- Zinn, J. S., Lavizzo-Mourey, R., & Taylor, L. (1993). Measuring satisfaction with care in the nursing home setting: The Nursing Home Resident Satisfaction Scale. *Journal of Applied Gerontology, 12*, 452-465.