

The Mental Health Statistics Improvement Program (MHSIP) Adult Consumer Satisfaction Survey Factor Structure and Relation to External Criteria

Alan Shafer, PhD

Rosemary Ang, MS

Abstract

The construct validity and generalizability of the Mental Health Statistics Improvement Program (MHSIP) adult consumer satisfaction survey was examined using a representative sample (N= 4242) of public mental health client surveys and assessment data. Factor analysis found five factors generally supporting the hypothesized scales of Satisfaction, Outcomes-Functioning, Access, Quality, and Social Connectedness that had acceptable reliability (average coefficient alpha = 0.91). The relation of the satisfaction survey scales to clinician-rated symptom scales, hospitalization, and functioning for 3 years prior to and after the survey was examined. Correlations were modest in magnitude but were in the predicted directions, with greater patient satisfaction being associated with lower symptoms and more positive outcomes. The largest predictors of the clinician-rated symptom scales were Outcomes-Functioning and Social Connectedness.

Introduction

In 1996, when the Center for Mental Health Services and the Substance Abuse and Mental Health Services Administration (SAMHSA) began funding the Mental Health Statistics Improvement Program (MHSIP), one of the required tasks was to develop a consumer-oriented report card of mental health services that would establish national performance benchmarks.¹ A task force of mental health agency representatives, clients, and advocates worked with the National Association of State Mental Health Program Directors to develop an adult consumer satisfaction

Address correspondence to Alan Shafer, PhD, Texas Health and Human Services Commission, Austin, TX, USA. .
Rosemary Ang, MS, Texas Health and Human Services Commission, Austin, TX, USA.

The MHSIP survey is conducted annually by the authors as a part of their job duties. This article was prepared by the authors in their own personal capacity. The opinions expressed are the authors' own and do not reflect any official policy, position, or view of any Texas Health and Human Service agency. The authors received no funding for this work and have no financial interests to declare.

Journal of Behavioral Health Services & Research, 2018, 565–578. © 2018 National Council for Behavioral Health. DOI 10.1007/s11414-018-9607-x

survey as a source for many of the report card performance measures. The task force identified specific elements of mental health services essential to consumer satisfaction: the access to services, the quality and appropriateness of services, and the outcomes of services, in addition to a global measure of satisfaction.² These became the MHSIP consumer satisfaction survey's main domains that were operationally defined as scales composed of individual items.

Access was defined as the quickness, ease, and convenience with which consumers received services. *Quality-Appropriateness* was defined as the quality and appropriateness of services the consumer received and how well the services matched his or her individual needs. Initially, the task force had two items defining a *Participation* domain indicating the clients' participation in treatment, but these items are highly similar to the Quality-Appropriateness domain and have been generally included in that scale. *Outcomes* was defined as the consumer's subjective perception of symptom reduction and improved functioning. *Satisfaction* was defined as how well the consumers liked the services overall. The initial items for these scales were derived from previous satisfaction surveys or were developed by the task force. Consumers rate each of the survey items on standard five-point Likert-type rating scales ranging from strongly agree to strongly disagree.

Previous factor analyses

Lebow reviewed factor analyses of a number of mental health satisfaction scales.^{3,4} Factor analysis of most of these scales resulted in a large general satisfaction factor. However, narrow delimited factors were also frequently found, which is consistent with a hierarchical factor structure of narrow lower-order factors and a general higher-order satisfaction factor. Frequent problems noted were that reliability was rarely reported, unrepresentative client samples were used, and surveys often had poor response rates.

Four previous studies have factor analyzed the MHSIP consumer satisfaction survey. A number of different versions of the consumer satisfaction survey have been developed which contain 20, 28, 36, or 40 items.^{1,2,5-8} Because of changes over time to the items in the consumer satisfaction surveys, the previous factor analyses of the survey rarely examine the same exact set of items twice. However, the overlap in items was generally high because the same pool of items was used for the consumer satisfaction surveys in the different studies, and the longer versions of the consumer satisfaction surveys typically contain all the items in the shorter versions of the consumer satisfaction surveys.

Wackwitz conducted factor analyses of the original 40-item consumer satisfaction survey to examine the psychometric characteristics and factor structure in order to explore the feasibility of creating a shorter version of the survey.⁵ The original 40-item consumer satisfaction survey produced clear factors for all four scales with reliability ranging from 0.81 to 0.94. A shorter 21-item survey was derived to measure the same four scales.

Eisen and colleagues also conducted a factor analysis of the original 40-item consumer satisfaction survey excluding the three Satisfaction items.⁶ A three-factor oblique solution clearly identified factors corresponding to the scales of Access, Quality-Appropriateness, and Outcomes with reliability ranging from 0.81 to 0.91.

Lutterman et al. conducted factor analyses of the 28-item consumer satisfaction survey in the process of developing the Social Connectedness and Functioning scales, finding a seven-factor solution with strong support for Quality-Appropriateness and the new scales of Social Connectedness and Functioning.⁷ Based in part on the results of this study, the final scales for Functioning and Social Connectedness were selected.

Jerrell conducted a factor analysis of the 20-item consumer satisfaction survey and found three factors that defined Access, Quality-Appropriateness, and Outcomes with an average reliability as measured by Cronbach's alpha of 0.77 and 2-week test-retest correlations of 0.55.⁸

The current study

In this study, the MHSIP consumer satisfaction surveys administered by the Texas Department of State Health Services, Mental Health and Substance Abuse division from 2004 to 2012 were

factor analyzed to determine the degree to which the original scales could be identified and the relation of the survey scales to external criteria. These annual surveys provide large, diverse random samples of adult public mental health service recipients diagnosed with a severe persistent mental illness.

The consumer satisfaction surveys are used to evaluate the performance of state mental health services and are required reporting for the SAMHSA's Community Mental Health Services Block Grant. However, Kaufmann and Phillips found that only 14% of the states administering these consumer satisfaction surveys had information about the reliability of their surveys and that less than 7% had conducted validity studies such as factor analysis.⁹ Approximately 400 to 500 million dollars are spent on the federal community mental health block grant annually, and since the consumer satisfaction survey is used for block grant reporting, it could be economically consequential if its results were used to help plan block grant spending priorities. Therefore, it is important to document the reliability and validity of the survey and to know the relation of the survey scales to other aspects of the treatment process.

Three previous factor analyses have called for additional research examining the relation of the consumer satisfaction survey to clinical symptoms, criteria validity separate from the survey, and analyzing the utility of the Social Connectedness scale.⁶⁻⁸ The purpose of this study was to examine the factor structure and reliability of the MHSIP consumer satisfaction surveys and to assess the concurrent validity of the scales with data external to the survey, as suggested by previous research. The relation of the survey scales to external criteria including symptom scales (as suggested by Jerrell and Shafer and Temple) and outcome measures such as hospitalization, substance abuse, housing, criminal justice involvement, and employment was used to examine the scales' concurrent validity.^{8,10} Clients indicating improved outcomes in the consumer satisfaction surveys should also exhibit improvements in clinical symptom scales and employment as well as decreased criminal justice involvement and hospitalization.

Method

Survey

A 28-item core survey was used annually during 2004 and 2005. The number of questions in each scale was Satisfaction [three], Outcomes [eight], Quality-Appropriateness [nine], Participation [two], and Access [six]. In 2006, eight additional items for Functioning [four] and Social Connectedness [four] were added to the survey. In this survey, lower scores on a one to five Likert-scale indicated greater satisfaction.

Sample

The 2004 to 2012 consumer satisfaction survey samples consisted of 4606 returned surveys, and 4242 were retained for analysis. Overall, 76% of the respondents had complete surveys, 14% were missing a single item, 2% were missing two items, and 8% were missing three or more items. Surveys retained for analysis could have no more than two missing items (92% of all clients). Surveys retained for analysis had missing items replaced with the mean.

There were 2590 (61%) females and 1652 (39%) males in the sample. The racial-ethnic distribution of the clients in the sample was 2338 (55%) Whites, 889 (21%) Hispanics, 853 (20%) Blacks, and 162 (4%) of some other race or ethnicity. The average age of clients in the sample was 46.7 (SD = 11.8). The most frequent primary diagnoses of clients were schizophrenia (34%), major depression (31%), and bipolar disorder (29%). A new random sample was drawn every year for the annual surveys. Clients may have participated in the surveys during multiple years; 237 (5%) clients had 2 years of surveys and 17 (<1%) had three surveys.

Procedure

The surveys used in this study were nine separate annual single statewide random samples (2004 to 2012) designed to be analyzed at the state level with a 95% confidence interval. Surveys were mailed out directly from the state headquarters to the clients' last known address. Clients included in the sample had received services from their local community mental health clinic during the survey year. On average, 512 surveys (SD = 76) were returned each year with an average survey raw return rate of 21% (SD = 6%).

Outcome data

In addition to the survey data, client records from mental health service databases were extracted in order to examine the relation of the survey to external criteria. The external criteria data for a client during the actual survey year itself was excluded, because the date surveys were conducted varied across the years and the date that a specific client's survey was returned was not available in the administrative database that records were obtained from. Data was extracted for each client for 3 years prior to their survey year and 3 years after the survey year. Data was then recoded from specific calendar years into years relative to the survey ranging from 3 years prior to 3 years after the survey for use in the analyses.

The clients' mental health service records included a variety of external criteria that were available for analysis. Symptom scales based on diagnoses (the Quick Inventory for Depressive Symptomatology (QIDS), the Positive and Negative Symptom Scales, and the Bipolar Depression Symptom Scale (BDSS)) were included. Records of state mental health hospitalizations were included in the study; however, hospitalizations in non-state hospitals were not available for inclusion. Additionally, a number of locally developed single-item rating scales for employment, housing, and other areas were included.

For clients diagnosed with depression, the 16-item self-report or clinician-rated QIDS scale was used.¹¹ The QIDS was derived from the longer Inventory of Depressive Symptomatology (IDS). The QIDS assesses all symptom domains needed to diagnosis depression. For clients diagnosed with bipolar disorder, the 10-item clinician-rated Brief Bipolar Disorder Symptom Scale (BDSS) was used. The BDSS was derived from the 24-item Brief Psychiatric Rating Scale (BPRS) to provide a short reliable measure of symptom severity.¹² For clients diagnosed with schizophrenia, two scales, one measuring positive symptoms and one measuring negative symptoms, were used. The four-item Brief Negative Symptom Assessment (BNSA) instrument is a clinician rating scale that is derived from the Negative Symptom Assessment and designed to determine the severity of negative symptoms of schizophrenia.¹³ Similarly, the four-item Positive Symptom Rating Scale (PSRS) is a locally developed clinician rating scale that is derived from the BPRS and designed to determine the severity of positive symptoms of schizophrenia.¹⁴ The PSRS consists of four BPRS items which have been found in meta-analyses of the 18-item BPRS and 24-item BPRS to be good indicators of positive symptoms, including suspiciousness, unusual thought content, hallucinations, and conceptual disorganization.^{15,16} The clinicians completing these scales were all licensed mental health professionals.

The Texas Recommended Assessment Guidelines (TRAG) are single-item clinician rating scales, with scores ranging from one (least severe) to five (most severe).¹⁷ Scales are anchored with specific behaviors and events to measure the areas of: social support, substance abuse, medical treatment, housing instability, harm to self or others, functional impairment, employment problems, and criminal justice involvement.

Because clients could potentially have many symptom scales and TRAG assessments on record during any particular service year, an average was computed for the multiple assessments in each individual year prior to and after the survey. Additionally, since data from the survey year was

excluded, analyses were conducted using data for clients that occurred either prior to or after the survey, which the majority of clients had. For example, a client may have had services the year before and after the survey but in no other years, or they may have had services and data only for the year prior to the survey and no data after the survey. In the correlational analyses of outcome data with the satisfaction survey, pairwise correlations were used to control for missing data. For state mental health hospitalization, a categorical variable indicating any hospitalization was used for analysis. Although some clients may have died or otherwise been lost to follow-up, on average, the clients had almost 2 years of assessments in the years after their survey ($M = 1.88$), so the number potentially lost to follow-up was relatively small.

Results

Factor analysis and scale reliability of the consumer satisfaction survey

Exploratory, rather than confirmatory, factor analysis was used in this study because confirmatory factor analysis is limited in that it only tests the specific items and factors that are specified. Because of the variability in previous studies, exploratory factor analysis allowed the factors and the items defining them to emerge and be examined without constraints. While previous studies have generally found support for most of the overall hypothesized structure of the survey scales, there has been enough variability both in the failure of factors to emerge and in the details of the items that define the factors to warrant further exploratory analyses.

Five factors were extracted using principal factors with squared multiple correlations as initial estimates of the communalities. Factors were rotated using direct oblique minimization. The first ten eigenvalues were 15.73, 3.94, 0.98, 0.73, 0.34, 0.26, 0.21, 0.18, 0.15, and 0.10. Given the design of the survey as five to seven conceptual domains and that the differences in eigenvalues after the fifth factor alternate between 0.03 and 0.05, a five-factor solution was used. The five factors were the closest possible solution to the original domains, and the resulting pattern of factor loadings was very clear with virtually no secondary factor loadings. The five-factor solution accounted for 60.3% of the variance. The results of the factor analyses are presented in Table 1.

Factor one was defined by the eight Outcomes items and the four Functioning items (average factor loading = 0.73). Factor two was defined by 11 items, the nine Quality-Appropriateness items, and the two Participation items (average loading = 0.63). Factor three was defined by the six Access items (average loading = 0.61). Factor four was defined by the four Social Connectedness items (average loading = 0.72). Factor five was defined by the three Satisfaction items (average loading = 0.64). Across all factors, the items defining them had large univocal factor loadings with no secondary salient loadings.

The factor analysis results described previously were for the 3282 clients whose surveys contained all 36 items from 2006 to 2012. Surveys in 2003 and 2004 did not contain the Functioning or Social Connectedness items. However, the results of a factor analysis of the 28 items common across all years revealed that the four resulting factors of Quality-Participation, Access, Outcomes, and Satisfaction were almost exactly the same as in the five-factor solution described previously. Details are available from the authors.

The total score for the satisfaction survey was 57.4 ($SD = 19.9$). Temporal trends in the total scale scores were examined by year, while an ANOVA revealed a significant difference among the annual mean scores. Scheffe's post hoc test indicated that no specific years were statistically significantly different from one another.

Scales were constructed based on the factor analysis results and the conceptual scale domains. Descriptive statistics for the total of the raw scale scores and for the average scale score item (based on the five-point Likert-type question format) were calculated. Scale reliability was assessed both by coefficient alpha for items at a particular point in time as well as test-retest correlations for 125 pairs of surveys completed by the same clients across two survey sample years (90 pairs for

Table 1

Exploratory factor analysis of satisfaction items. Direct oblimin rotated factor pattern (standardized regression coefficients)

Scale	Item	F1	F2	F3	F4	F5	HSQ
Q01	I like the services I received here. (S)	0.08	0.04	0.22	0.02	0.63	0.749
Q02	If I had other choices, I would still get services from this agency. (S)	0.04	0.07	0.09	0.04	0.62	0.594
Q03	I would recommend this agency to a friend or family member. (S)	0.04	0.13	0.06	0.03	0.69	0.724
Q04	The location of services was convenient (parking, public transportation, distance, etc.). (A)	0.00	0.02	0.34	0.03	0.23	0.314
Q05	Staff were willing to see me as often as I felt it was necessary. (A)	0.02	0.02	0.74	-0.01	0.09	0.671
Q06	Staff returned my call in 24 h. (A)	0.01	0.11	0.68	-0.01	-0.05	0.541
Q07	Services were available at times that were good for me. (A)	0.01	0.05	0.64	0.01	0.14	0.644
Q08	I was able to get all the services I thought I needed. (A)	0.05	0.05	0.59	0.05	0.19	0.667
Q09	I was able to see a psychiatrist when I wanted to. (A)	0.02	0.05	0.70	0.03	-0.01	0.577
Q10	Staff here believe that I can grow, change and recover. (Q)	0.21	0.49	0.04	-0.01	0.09	0.511
Q11	I felt comfortable asking questions about my treatment and medication. (Q and P)	0.05	0.52	0.10	0.04	0.18	0.614
Q12	I felt free to complain. (Q)	0.00	0.45	0.21	0.05	0.10	0.524
Q13	I was given information about my rights. (Q)	-0.06	0.71	-0.04	0.03	0.10	0.530
Q14	Staff encouraged me to take responsibility for how I live my life. (Q)	0.03	0.78	-0.04	0.03	0.01	0.611
Q15	Staff told me what side effects to watch out for. (Q)	-0.01	0.69	0.13	0.02	-0.06	0.572
Q16	Staff respected my wishes about who is and who is not to be given information about my treatment. (Q)	-0.09	0.76	-0.04	0.05	0.06	0.566
Q17	I, not staff, decided my treatment goals. (Q and P)	0.17	0.56	0.02	-0.02	-0.02	0.434
Q18	Staff were sensitive to my cultural background (race, religion, language, etc.) (Q)	-0.01	0.68	0.02	0.02	0.03	0.513
Q19	Staff helped me obtain the information I needed so that I could take charge of managing my illness. (Q)	0.12	0.67	0.10	-0.04	0.02	0.668
Q20	I was encouraged to use consumer-run programs (support	0.04	0.61	0.11	0.00	-0.08	0.464

Table 1
(continued)

Scale	Item	F1	F2	F3	F4	F5	HSQ
	groups, drop-in centers, crisis phone line, etc.). (Q)						
Q21	I deal more effectively with daily problems. (O)	<i>0.74</i>	0.13	-0.01	-0.08	0.07	0.623
Q22	I am better able to control my life. (O)	<i>0.89</i>	0.08	-0.07	-0.08	0.08	0.759
Q23	I am better able to deal with crisis. (O)	<i>0.91</i>	0.01	-0.01	-0.09	0.05	0.760
Q24	I am getting along better with my family. (O)	<i>0.49</i>	0.00	0.11	0.23	0.03	0.535
Q25	I do better in social situations. (O)	<i>0.73</i>	0.05	0.05	0.09	-0.07	0.669
Q26	I do better in school and/or work. (O)	<i>0.66</i>	0.03	0.03	0.04	-0.07	0.487
Q21	I deal more effectively with daily problems. (O)	<i>0.43</i>	0.04	0.11	0.20	-0.01	0.452
Q22	I am better able to control my life. (O)	<i>0.77</i>	-0.02	0.03	0.05	-0.01	0.648
Q29	I do things that are more meaningful to me.(F)	<i>0.65</i>	0.00	0.03	0.18	0.01	0.645
Q30	I am better able to take care of my needs. (F)	<i>0.81</i>	-0.05	0.01	0.06	0.05	0.709
Q31	I am better able to handle things when they go wrong. (F)	<i>0.90</i>	-0.01	-0.01	-0.02	-0.01	0.764
Q32	I am better able to do things that I want to do. (F)	<i>0.76</i>	-0.06	0.03	0.14	0.01	0.715
Q33	I am happy with the friendships I have. (SC)	0.06	0.02	0.00	<i>0.77</i>	0.02	0.681
Q34	I have people with whom I can do enjoyable things. (SC)	0.01	0.07	-0.10	<i>0.84</i>	0.01	0.701
Q35	I feel I belong in my community.(SC)	0.19	-0.06	0.11	<i>0.62</i>	-0.03	0.604
Q36	In a crisis, I would have the support I need from family or friends. (SC)	-0.03	0.07	0.05	<i>0.65</i>	0.03	0.479
	Inter-factor correlations	F1	F2	F3	F4	F5	
F1	Satisfaction	1.00	0.51	0.49	0.67	0.32	
F2	Access		1.00	0.80	0.40	0.66	
F3	Quality			1.00	0.40	0.72	
F4	Outcomes-Functioning				1.00	0.22	
F5	Social Connectedness					1.00	

Original domain scale indicated by letter in the parenthesis at the end of each item. Highest factor loading in italics
O outcome, *S* satisfaction, *Q* quality-appropriateness, *F* functioning, *A* access, *P* participation, *SC* social connectedness, *HSQ* factor analysis communality

Outcomes-Functioning and Social Connectedness), with an average time between surveys of 2.74 years (SD = 1.75). The descriptive statistics for the scales are summarized in Table 2.

Relation of the survey scales to outcome measures

The descriptive statistics for the external criteria (symptom and outcome measures) are displayed in Table 3. The means are arranged from 3 years before the satisfaction survey to 3 years after the

Table 2

Reliability estimates for satisfaction factor-based scales

Scale	Raw	Raw	Five point	Five point	No.	Coefficient	Test
	Mean	SD	Mean	SD		Alpha	
Satisfaction	5.24	2.49	1.75	0.83	3	0.88	0.39
Access	11.75	4.99	1.96	0.83	6	0.88	0.38
Quality	21.22	8.20	1.93	0.75	11	0.93	0.51
Outcomes*	19.03	7.14	2.38	0.89	8	0.93	0.62
Outcomes-functioning*	28.76	10.71	2.40	0.89	12	0.95	0.63
Social connectedness	8.93	3.78	2.23	0.94	4	0.87	0.62
AVERAGE						0.91	0.53

The outcome scale consists of eight items administered through 2006, functioning adds four additional items to outcomes starting in 2007. Because they load on the same factor, they form a single larger scale from 2007 onwards. There were 125 test-retest pairs of surveys (90 pairs for outcomes-functioning and social connectedness) with an average time between surveys of 2.74 years (SD = 1.75)

survey. Because a client could have multiple assessments during any particular year (or none if they were not in service), the means in each column were calculated using the average score for each client in a particular year (and then averaged across all clients normally). Some clients may have had multiple assessments in a year and some clients may have only a single one. The vast majority (95%) of satisfaction survey clients had external criteria data in the 3 years prior to or after the survey, and, on average, the clients had 3.96 years of external criteria data (SD = 1.48) available for analysis. The general trend for the external criteria data was from higher scores (indicating more difficulties) prior to the satisfaction survey to lower scores (indicating improvement) after the survey.

However, the external criterion scores in Table 3 are for group means, and thus, the correlations that are displayed in Table 4 should be used to examine the relation of the satisfaction survey scales with outcomes and symptoms. The correlations are ordered from 3 years before to 3 years after the satisfaction survey and in subsets for each of the satisfaction scales. For these results, higher symptom scale scores, hospitalization, or higher TRAG scores indicate worse outcomes, and high satisfaction scale scores indicate lower satisfaction (the satisfaction items are scored from one = strongly agree to five = strongly disagree) so positive correlations indicate that low satisfaction is associated with negative outcomes and that high satisfaction is associated with positive outcomes.

For the survey scales of Access, Quality-Participation, and Satisfaction, there were relatively few significant correlations, primarily with the outcome scales for depression and bipolar disorders (QIDS and BDSS). However, for the Outcomes-Functioning scale, there were a large number of significant correlations of larger magnitude than the three previous scales, likely because the content of the Outcomes and Functioning scale items is more similar to the external criteria measures for symptoms and outcomes. All four symptom scales had relatively large significant correlations across all the years. The Social Connectedness scale had a number of relatively large and significant correlations with the external criteria measures. All four symptom scales had relatively large significant correlations with Social Connectedness across all the years. The TRAG scales also had significant correlations across many years.

Follow-up regression analyses were conducted fitting a general model of the five satisfaction scales (Satisfaction, Access, Quality, Outcomes-Functioning, and Social Connectedness) as predictors of the individual symptom scales 3 years prior to and after the satisfaction survey. The results are summarized here, but the detailed results are available from the authors. Overall, all of the regressions were statistically significant. However, the average adjusted *R* square was in the single digits, indicating a relatively small

Table 3

Outcome item descriptive statistics relative to years prior and after the clients' satisfaction survey

	Mean <i>N</i>	Years before 3	Years before 2	Years before 1	Years after 1	Years after 2	Years after 3
Any hospital discharge (<i>N</i>)	4242	145	184	196	88	60	53
Percent		3.4%	4.3%	4.6%	2.1%	1.4%	1.2%
BDSS (bipolar disorder)	872	20.54	21.27	21.42	19.12	18.87	18.73
SD		7.54	7.94	8.04	7.00	7.09	7.07
QIDS (depression)	803	11.57	11.53	11.70	9.94	9.42	9.14
SD		5.42	5.24	5.29	4.85	4.82	4.72
BNSA (negative symptoms)	1040	8.20	8.32	8.30	7.66	7.58	7.71
SD		3.45	3.57	3.48	3.22	3.20	3.30
PSRS (positive symptoms)	1041	8.68	8.70	8.72	7.93	7.76	7.88
SD		4.31	4.41	4.29	3.90	3.77	3.95
TRAG support	2628	2.32	2.35	2.35	2.18	2.13	2.13
SD		0.81	0.82	0.81	0.77	0.77	0.77
TRAG substance abuse	2629	1.21	1.20	1.21	1.14	1.13	1.12
SD		0.55	0.51	0.52	0.40	0.38	0.41
TRAG medical treatment	858	1.56	1.54	1.59	1.44	1.39	1.38
SD		0.69	0.68	0.69	0.61	0.59	0.58
TRAG housing instability	2628	1.77	1.81	1.79	1.60	1.56	1.52
SD		0.86	0.86	0.85	0.75	0.74	0.70
TRAG harm to self or others	2631	1.44	1.45	1.45	1.29	1.28	1.30
SD		0.62	0.65	0.64	0.52	0.50	0.57
TRAG functional impairment	2630	2.43	2.47	2.46	2.26	2.23	2.26
SD		0.80	0.81	0.80	0.75	0.75	0.75
TRAG Employment problems	2628	1.96	1.95	1.90	1.50	1.38	1.33
SD		1.37	1.37	1.32	1.05	0.92	0.84
TRAG criminal justice	2628	1.23	1.22	1.25	1.18	1.15	1.13
SD		0.74	0.69	0.75	0.63	0.59	0.55

magnitude for the overall effect size. The most interesting results were first, that the Outcomes-Functioning satisfaction scale was often (but not always) the largest significant predictor. Second, the Social Connectedness scale was also equally often a significant predictor with only a somewhat smaller magnitude than Outcomes-Functioning. The scales for Access and Quality were significant predictors slightly less than half as many times as the scales for Social Connectedness and Outcomes-Functioning. The Satisfaction scale itself was only a significant predictor once. Interestingly, the average *R* square for predicting the scales for depression symptoms (0.09) was larger than for bipolar disorder symptoms (0.06), which was larger than both positive and negative schizophrenia symptoms (0.03).

Discussion

The majority of the original hypothesized domain scales for the satisfaction surveys formed well-defined factors, and scales based on the items defining these factors had high internal

Table 4

Correlations of satisfaction scales with outcome items relative to years prior and after the client's satisfaction survey

	Years before 3	Years before 2	Years before 1	Years after 1	Years after 2	Years after 3
SATISFACTION						
Any hospitalization*	-0.00	0.03	0.03	0.03	0.02	-0.01
BDSS	<i>0.08</i>	<i>0.10</i>	0.07	<i>0.12</i>	<i>0.09</i>	0.03
QIDS	<i>0.13</i>	<i>0.09</i>	<i>0.09</i>	<i>0.10</i>	<i>0.13</i>	<i>0.10</i>
BNSA (negative symptoms)	<i>0.08</i>	0.03	0.03	0.01	-0.01	0.00
PSRS (positive symptoms)	0.03	0.02	0.04	0.04	0.01	0.00
TRAG support	0.06	0.05	0.05	0.03	0.07	0.05
TRAG substance abuse	0.07	0.05	0.05	0.04	0.04	0.07
TRAG medical treatment	<i>0.09</i>	0.05	0.02	0.06	<i>0.08</i>	0.06
TRAG housing instability	0.06	0.06	0.05	0.03	0.05	0.06
TRAG harm to self or others	0.04	0.06	0.03	0.05	0.05	0.05
TRAG functional impairment	<i>0.08</i>	0.06	0.03	0.06	0.06	0.04
TRAG employment problems	0.06	0.05	0.02	0.02	0.06	0.05
TRAG criminal justice	0.05	0.04	0.02	0.04	0.04	0.05
ACCESS						
Any hospitalization*	-0.02	0.03	0.01	0.02	0.01	-0.03
BDSS	<i>0.08</i>	<i>0.14</i>	<i>0.10</i>	<i>0.15</i>	<i>0.14</i>	<i>0.09</i>
QIDS	<i>0.20</i>	<i>0.15</i>	<i>0.13</i>	<i>0.13</i>	<i>0.10</i>	<i>0.09</i>
BNSA (negative symptoms)	0.07	0.03	0.04	0.05	0.03	0.04
PSRS (positive symptoms)	0.03	0.04	0.07	0.07	0.06	0.04
TRAG support	<i>0.08</i>	0.06	0.07	0.06	<i>0.08</i>	0.05
TRAG substance abuse	<i>0.08</i>	0.07	<i>0.08</i>	0.06	<i>0.08</i>	<i>0.09</i>
TRAG medical treatment	<i>0.13</i>	<i>0.09</i>	0.05	0.05	0.06	0.01
TRAG housing instability	<i>0.09</i>	<i>0.08</i>	<i>0.08</i>	0.06	0.07	<i>0.08</i>
TRAG harm to self or others	0.05	0.07	0.05	0.08	0.08	0.05
TRAG functional impairment	0.07	0.07	0.05	<i>0.09</i>	0.07	0.05
TRAG employment problems	<i>0.08</i>	0.06	0.06	0.05	0.06	0.05
TRAG criminal justice	<i>0.08</i>	0.06	0.04	0.07	0.05	0.06
QUALITY						
Any hospitalization*	-0.02	0.01	0.01	0.02	0.01	-0.03

Table 4
(continued)

	Years before 3	Years before 2	Years before 1	Years after 1	Years after 2	Years after 3
BDSS	0.03	0.12	0.09	0.10	0.09	0.07
QIDS	0.14	0.12	0.10	0.10	0.10	0.09
BNSA (negative symptoms)	0.06	0.04	0.03	0.02	0.00	0.00
PSRS (positive symptoms)	0.03	0.03	0.05	0.04	0.00	0.02
TRAG support	0.05	0.04	0.04	0.04	0.05	0.06
TRAG substance abuse	0.04	0.05	0.04	0.03	0.03	0.05
TRAG medical treatment	0.09	0.07	0.04	0.04	0.06	0.02
TRAG housing instability	0.07	0.05	0.05	0.01	0.03	0.05
TRAG harm to self or others	0.00	0.03	0.02	0.06	0.05	0.03
TRAG functional impairment	0.06	0.05	0.02	0.06	0.05	0.05
TRAG employment problems	0.06	0.05	0.03	0.03	0.04	0.04
TRAG criminal justice	0.08	0.04	0.02	0.04	0.03	0.04
OUTCOMES-FUNCTIONING						
Any hospitalization*	-0.04	-0.02	-0.01	-0.01	-0.01	-0.03
BDSS	0.23	0.22	0.19	0.24	0.21	0.18
QIDS	0.28	0.26	0.24	0.28	0.26	0.27
BNSA (negative symptoms)	0.19	0.18	0.16	0.13	0.11	0.11
PSRS (positive symptoms)	0.14	0.15	0.19	0.18	0.10	0.09
TRAG support	0.10	0.09	0.11	0.09	0.09	0.09
TRAG substance abuse	0.09	0.06	0.08	0.09	0.07	0.07
TRAG medical treatment	0.20	0.15	0.14	0.15	0.13	0.15
TRAG housing instability	0.10	0.09	0.11	0.05	0.01	0.03
TRAG harm to self or others	0.07	0.09	0.12	0.12	0.11	0.12
TRAG functional impairment	0.14	0.13	0.13	0.15	0.10	0.09
TRAG employment problems	0.11	0.09	0.08	0.06	0.05	0.04
TRAG criminal justice	0.06	0.06	0.05	0.06	0.05	0.05
SOCIAL CONNECTEDNESS						
Any hospitalization*	-0.04	-0.02	-0.01	-0.01	-0.02	-0.05
BDSS	0.20	0.18	0.19	0.20	0.14	0.13
QIDS	0.27	0.25	0.20	0.21	0.25	0.26
BNSA (negative symptoms)	0.19	0.16	0.16	0.11	0.11	0.12
PSRS (positive symptoms)	0.16	0.15	0.16	0.15	0.11	0.09
TRAG support	0.16	0.14	0.14	0.12	0.12	0.11
TRAG substance abuse	0.08	0.06	0.06	0.07	0.08	0.07
TRAG medical treatment	0.19	0.12	0.11	0.11	0.17	0.19
TRAG housing instability	0.11	0.11	0.11	0.06	0.02	0.06
TRAG harm to self or others	0.11	0.11	0.11	0.10	0.09	0.08
TRAG functional impairment	0.16	0.15	0.11	0.12	0.09	0.07

Table 4
(continued)

	Years before 3	Years before 2	Years before 1	Years after 1	Years after 2	Years after 3
TRAG employment problems	<i>0.10</i>	<i>0.10</i>	<i>0.10</i>	0.06	0.04	0.05
TRAG criminal justice	0.05	0.06	0.04	0.05	0.04	0.05

*Any hospitalization used a point biserial correlation for the categorical outcome of any hospitalization
Italicized correlations are significant at < 0.0001 and have an absolute magnitude > 0.08 . Positive correlations are good, because higher symptom, hospital, or TRAG scores indicate worse outcomes and higher satisfaction scale scores indicate lower satisfaction. So as satisfaction scales decrease, negative outcomes increase (e.g., high negative outcomes)

consistency (average coefficient alpha = 0.91), and for a small subsample, reasonable test-retest correlations ($r = .53$) across an average of 2.7 years. The results of this study, in conjunction with previous research, support the structure and reliability of most of the scales. The satisfaction survey scales also had a consistent pattern of modest correlations with the external criteria of symptom scales and other outcomes such that lower client satisfaction was associated with higher symptoms and poorer outcomes, and higher client satisfaction was associated with lower symptoms and better outcomes, which also provides evidence for the satisfaction surveys construct validity and generalizability.

Similar to the findings of Wackwitz, Jerrell, and Eisen et al., the results of the factor analysis in this study clearly support the Access, Quality, and Outcomes factors.^{5,6,8} Also like Lutterman et al., the Social Connectedness scale was supported as an independent factor.⁷ The results from the present study indicated that the Functioning items were not clearly distinguishable from the Outcomes items, similar to the results found by Shafer and Temple for the youth and family version of these scales, which suggests they should be combined into a single scale.¹⁰ The two Participation items were not distinguishable from the Quality-Appropriateness items, and they should also probably be combined into a single scale.

The pattern of the client's satisfaction scale correlations to outcomes was as predicted with lower satisfaction scores being associated with worse outcomes and higher satisfaction scores being associated with better outcomes. The specific satisfaction scales of Outcomes-Functioning and Social Connectedness were the most strongly associated with outcomes. The Social Connectedness scale is especially interesting and important, because this domain is one of the SAMHSA's national outcome measures and has items reflecting both social support and loneliness, both of which are likely to have an effect on mental health. The symptom scales for bipolar depression and major depression also generally had higher correlations with satisfaction than did the positive and negative symptom scales for schizophrenia. The majority of external criteria were completed by clinicians (with the exception of approximately half the QIDS that can be self-report or clinician rated). This pattern of correlations indicates that the clinician's ratings of symptoms and the client's perception of their own functioning and outcomes are systematically related in ways that would be expected. Overall, although generally modest, these correlations with external criteria contribute to documenting the construct validity and generalizability of these satisfaction scales. Additionally, the pattern of results across time was that symptoms and outcomes were generally worse before the satisfaction survey; yet, symptoms and outcomes were generally better after the satisfaction survey, possibly suggesting that the overall length of treatment may have a generally positive effect, with lower symptoms and more positive outcomes as the number of years of treatment increases.

The most objective outcome, psychiatric hospitalizations, did not have any large significant correlations with the survey scales. However, the administrative databases used included only hospitalizations in state hospitals and did not include other private or public hospitalizations and, therefore, limited the data available for analysis. In addition, psychiatric hospitalization is relatively rare, with only about 11% of the sample having any hospitalizations; hence, the base rate may have been too low to detect any effect.

This study had a number of limitations, one of which was that the survey response rate was relatively low. Another limitation was that the matching of the client's assessment data to satisfaction surveys required excluding external data for the actual year of the survey due to the lack of exact survey completion dates. Consequently, symptom scales and outcomes were further removed in time from the satisfaction survey and were only available for the years before and after the survey. In addition, the further away from the survey year one goes in time, the somewhat fewer symptom scales and outcomes were available for analysis. Both of these issues may have attenuated the correlations between the survey scales and the outcomes. Nevertheless, the sample was a large group of public mental health clients that typically had multiple years of data available for each client, which did have some significant but modest correlations with external criteria.

Implications for Behavioral Health

The MHSIP consumer satisfaction surveys are required of states by SAMHSA as part of the Community Mental Health Services Block Grant. The survey is used to evaluate the performance of state mental health services statewide, or by a provider, and is used to improve services. The surveys are necessary to provide the consumer's perspective on the quality and impact of services received. Based on a state's survey results, states will provide additional assistance to improve services and the client's satisfaction with those services, if needed. Given the significance attached to the results of the survey at the state and national levels, it is important to understand the strengths and limitations of the survey.

The factor analysis found strong support for the predicted satisfaction scale structure and scales based on these factors showed acceptable reliability. This finding supports the structure and reliability of the survey scales in conjunction with previous factor analysis. In this study, the clients' personal assessment of their outcomes and functioning was generally congruent with objective assessments of client symptoms by clinicians. Moreover, symptom improvement was predicted not only by client assessment of Outcomes and Functioning but also by Social Connectedness with small additional contributions from Quality and Appropriateness. This represents an important step forward in demonstrating the construct validity and generalizability of the consumer satisfaction survey as well as the potential utility of satisfaction surveys to assess client's treatment experiences and outcomes.

The Outcomes-Functioning scale is important, because while individual symptom data for clients is not collected at the national level, the survey scales are collected and reported by states. These results suggest that the clients' Outcomes-Functioning scale scores could be used as a reasonably valid (although modest) indicator of the clients' clinical outcomes. This might be of some assistance at the aggregate level for program evaluation and planning for the Community Mental Health Services Block Grant. Yet, more research on using this as an indicator of clinical outcomes would be needed to further establish its validity.

The Social Connectedness scale is unique and different from most satisfaction survey scales and probably serves as a measure of social support. This should be investigated further in future research as social support plays an important role in mediating many health outcome processes and may be an area that some clients could potentially strengthen to assist in their recovery given its unique independent contribution to symptom improvement.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflicts of interest.

References

1. TeagueGB, GanjuV, HornikJA, et al.The MHSIP Mental Health Report Card. A Consumer-Oriented Approach to Monitoring the Quality-Appropriateness of Mental Health Plans. Mental Health Statistics Improvement Program. *Evaluation Review*.1997;21(3):330–341.
2. GanjuV. *The MHSIP Consumer Survey: History, Development, Revisions, Applications, and Commonly Asked Questions*. Austin, TX: Texas Department of Mental Health and Mental Retardation, 1999.
3. LebowJL. Consumer Satisfaction with Mental Health Treatment. *Psychological Bulletin*.1982;91(2):244–259.
4. LebowJL. Research Assessing Consumer Satisfaction with Mental Health Treatment: A Review of Findings. *Evaluation and Program Planning*.1984;6(3–4):211–236.
5. WackwitzJH. *Consumer Survey Analysis: Memo to MHSIP Workgroup*. Denver, CO: Colorado Mental Health Services, 2000.
6. EisenSV, ShaulJA, LeffHS, et al.Toward a National Consumer Survey: Evaluation of the CABHS and MHSIP Instruments. *Journal of Behavioral Health Services & Research*.2001;28(3):347–369.
7. LuttermanT, PhelanBE, BerhaneA. *Development of Mental Health Measure of Social Connectedness*. Alexandria, VA: National Association of State Mental Health Program Directors, 2007.
8. JerrellJM. Psychometrics of the MHSIP Adult Consumer Survey. *Journal of Behavioral Health Services & Research*.2006;33(4):483–488.
9. KaufmannC, PhillipsD. *Survey of State Consumer Surveys*. Rockville, MD: Substance Abuse and Mental Health Services Administration, 2000.
10. ShaferA, TempleJ. Factor Structure of the Mental Health Statistics Improvement Program (MHSIP) Family and Youth Satisfaction Surveys. *Journal of Behavioral Health Services & Research*.2013;40(3):306–316.
11. RushAJ, TrivediMH, IbrahimHM, et al.The 16-item Quick Inventory of Depressive Symptomatology (QIDS) Clinician Rating (QIDS-C) and Self-Report (QIDS-SR): A Psychometric Evaluation in Patients with Chronic Major Depression. *Biological Psychiatry*.2003;54(5):573–583.
12. DennehyEB, SuppesT, CrismonML, et al.Development of the Brief Bipolar Disorder Symptom Scale for Patients with Bipolar Disorder. *Psychiatry Research*.2004;127(1–2):137–145.
13. AlphsL, MorlockR., CoonC, et al.The 4-Item Negative Symptom Assessment (NSA-4) Instrument: A Simple Tool for Evaluating Negative Symptoms in Schizophrenia Following Brief Training. *Psychiatry (Edgmont)*.2010;7(7):26–32.
14. OverallJE, GorhamDR. The Brief Psychiatric Rating Scale. *Psychological Reports*.1962;10(3):799–812.
15. ShaferA. Meta-Analysis of the Brief Psychiatric Rating Scale Factor Structure. *Psychological Assessment*.2005;17(3):324–335.
16. DazziF, ShaferA, Lauriola, M. Meta-Analysis of the Brief Psychiatric Rating Scale-Expanded (BPRS-E) Structure and Arguments for a New Version. *Journal of Psychiatric Research*.2016;81:140–151.
17. Texas Department of State Health Services. *User's Manual for the Adult Texas Recommended Assessment Guidelines (Adult-TRAG)*.Austin, TX: Mental Health and Substance Abuse Division, 2011.