



uSPEQ® Consumer Experience Survey Psychometric Evaluation

Prepared by:

uSPEQ Research
and Development Team

July 2008

uSPEQ®

4891 East Grant Road
Tucson, Arizona 85712 USA
Voice (888) 877-3788
Fax (888) 789-7367
Fax (520) 318-1129
www.uspeq.org

© 2008 by uSPEQ.

All rights reserved • Published 2008 • Printed in the United States of America

Any copying, republication, or redistribution of the content by any means is expressly prohibited. Unauthorized use of any content may violate copyright laws, trademark laws, the laws of privacy and publicity, and communications regulations and statutes. Data is provided for information purposes only and is not intended for trading purposes.

Table of Contents

Introduction	1
The Survey Instrument	3
Methods.....	7
Pilot Testing (uSPEQ v.1.0)	7
Rasch Modeling (Item Response Theory).....	8
Classical Test Theory Procedures.....	10
Survey Readability	13
Independent Survey Sample	13
uSPEQ Instrument Refinement (v.2.0)	13
Survey Distribution Methods	14
Survey Sample.....	14
Factor Analysis	14
Rasch Modeling.....	15
Classical Test Theory Procedures.....	15
Conclusion	19
Appendix A: Descriptive Statistics on Pilot Study Participants	21
Respondents by Age	21
Respondents by Race and Ethnicity.....	22
Respondents by Education.....	23
Respondents by Completion/Help Methods	24
Appendix B: Descriptive Statistics on Version 1.0 Participants	25
Respondents by Age	25
Respondents by Race and Ethnicity.....	26
Respondents by Education.....	27
Respondents by Completion/Help Methods	27
References	29

Table of Illustrations

Figure 1.	uSPEQ Questionnaire with Three Tiers of Survey Items	4
Table 1.	Distribution of Respondents by Provider Service Area	7
Table 2.	Summary Reliability Statistics by Factor	9
Table 3.	Cronbach’s Alpha Values for uSPEQ Domains.....	10
Figure 2.	Average Convergent and Divergent Validity	11
Table 4.	Multiple Regression F-Statistics for 7-Item Model	12
Table 5.	Multiple Regression F-Statistics for 49-Item Model.....	12
Table 6.	uSPEQ v.1.0: Distribution of Respondents by Provider Service Area	14
Table 7.	uSPEQ v.2.0: Summary Reliability Statistics by CSU.....	15
Table 8.	uSPEQ v.2.0: Reliability Statistics (Cronbach’s Alpha).....	16
Figure 3.	uSPEQ v.2.0: Average Convergent and Divergent Validity.....	16
Table 9.	uSPEQ v.2.0: Multiple Regression F-Statistics for 19-Item Model	17
Figure 4.	uSPEQ v.2.0: Tier 1/Tier 2 Modular Approach.....	19
Table 10.	Pilot Study: Respondents by Provider Service Area.....	21
Table 11.	Pilot Study: Respondents by Age	21
Table 12.	Pilot Study: Respondents by Race	22
Table 13.	Pilot Study: Respondents by Education Level	23
Table 14.	Pilot Study: Respondents by Completion/Help Methods	24
Table 15.	Version 1.0: Respondents by Provider Service Area.....	25
Table 16.	Version 1.0: Respondents by Age	25
Table 17.	Version 1.0: Respondents by Race	26
Table 18.	Version 1.0: Respondents by Education Level.....	27
Table 19.	Version 1.0: Study Respondents by Completion/Help Methods	27

Introduction

uSPEQ® (pronounced *you speak*) is a confidential, anonymous, and scientifically-tested consumer reporting system that gives persons served a voice in their services. The uSPEQ Consumer Experience Survey is a subjectively measured, self-administered instrument consisting of a 20-item questionnaire (version 2.0) with five domains or subscales. The primary purpose of uSPEQ is to gather feedback from consumers or persons served regarding their perceptions of the quality of care or services they are currently receiving or have received in the past. Providers across the spectrum of health and human services can use the uSPEQ feedback for quality improvement and outcomes management. As a uniform survey tool for all health and human services fields, uSPEQ has the ability to assess performance across diverse populations and settings. This survey tool is designed to provide data for benchmarking and comparative analysis of the consumer experience. The survey instrument was field tested to ensure its psychometrical soundness as well as feasibility for data collection in the fields. The questionnaire was recently further refined based on additional survey data collected. This brief report provides a summary of the development history and psychometric evaluations of uSPEQ Consumer Experience Survey.

In today's competitive markets, ensuring quality care is a primary concern for all service providers. As the cost of services continue to rise, consumers are better educated and are asking service providers to demonstrate value as it relates to safety, quality, and customer satisfaction. uSPEQ was conceptualized to support providers as they demonstrate and communicate the value of their programs and services to the persons receiving services and to the public.

uSPEQ was developed and is administered under the auspices of CARF, a leading international accrediting body in the areas of Aging Services; Behavioral Health; Child and Youth Services; Durable Medical Equipment, Prosthetics, Orthotics, and Supplies; Employment and Community Services; Medical Rehabilitation; and Opioid Treatment Programs. The CARF family of organizations currently accredits more than 5,000 providers at more than 18,000 locations in the United States, Canada, Western Europe, and South America. More than 6.5 million persons of all ages are served annually by providers of CARF-accredited programs and services.

CARF began its work on performance indicators in 1997, and published its first monograph, **Performance Indicators for Rehabilitation Programs.**¹² A series of leadership panels, a national invitational conference, consumer focus groups, advisory committees, and a work group helped CARF refine its direction in a heavily populated field of players already engaged in developing indicators and measures for performance improvement. Three recurring themes caught the attention of members of CARF's performance indicators project as they reviewed the literature and gathered input from CARF's stakeholders:

- Consumers share many common concerns about services they receive and outcomes they attained—access to services, respect and involvement, information, safety, services directed to their needs, and meaningful participation in their lives.
- Providers want a tool that crosses multiple populations and settings so they can efficiently and cost-effectively use their data system dollars.

- Consumers and providers alike want to be able to compare themselves to the norm. Consumers ask, **What happens here for people like me?** Providers ask, **How do we compare with other organizations?** and, **Are we improving?**

Using the advice of its input groups and advisory councils, CARF developed an instrument and information system to address these needs. uSPEQ features a confidential and anonymous questionnaire to be completed by consumers. The questionnaire and data set include items that capture characteristics of the respondents and information about their program participation and how they completed the questionnaire. The uSPEQ questionnaire also asks respondents to rate their experiences related to access to services, the service process, the way the program meets their needs, and their perception of the outcomes they attained. Developed over a decade with the input of diverse stakeholders, uSPEQ is unique in several respects:

- Survey items are consumer based; i.e., the survey questions were developed with broad input from the consumers, and they are worded from the perspectives of the persons served.
- Items are crosscutting in nature, ensuring that the survey can be efficiently administered across all components of a service continuum.
- Domains span the concerns of persons served.
- Questionnaires and reports are customized to the needs of an organization, its programs, and its populations served.
- Survey aligns with important national and international disability and rehabilitation frameworks.

uSPEQ is defined as **crosscutting** because the concerns reflected in the questionnaire items cross lines of population and organization settings. Subscribers can utilize uSPEQ within any service setting and with any population. Furthermore, uSPEQ is specifically designed to address the needs of individual consumers regardless of age group, gender, educational background, race, ethnicity, and socioeconomic status in order to accurately reflect the diverse populations served by providers. It is the voice of the consumer. The focus is on the person who received the services, and it answers the question: **What happens to people like me in your program?**

uSPEQ gathers consumers' experiences with programs, services, and providers via online or paper questionnaires. In turn, providers use the reported information to improve the quality of programs and services.

The Survey Instrument

It is critical that the resulting data from uSPEQ answer key questions, not only for providers in their quality improvement programs or conformance to accreditation standards, but also for the human service fields and CARF itself. A guiding principle was that uSPEQ should reflect the domains of concern consistent with key conceptual frameworks related to assessing and improving the lives of persons served, and to leveling the playing field for persons with disabilities. The domains, data elements, and questions for respondents are consistent with the following frameworks:

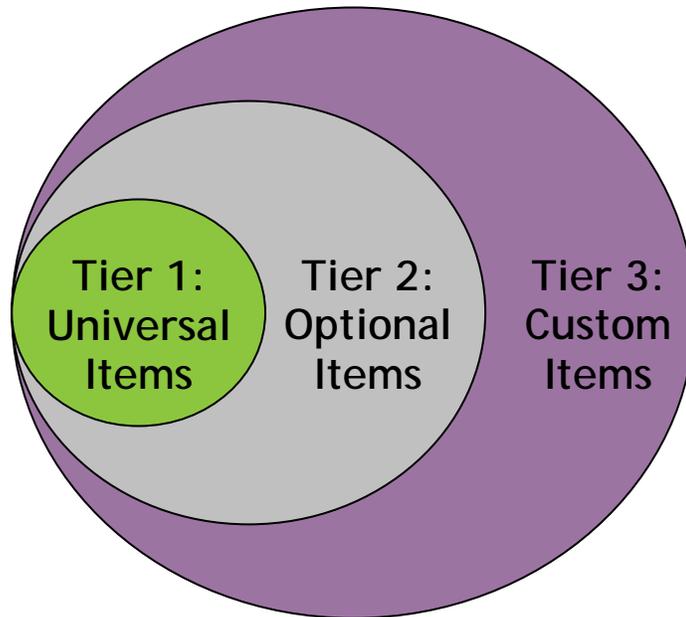
- **The World Health Organization’s (WHO) International Classification of Functioning Disability and Health.**¹³ The ICF framework is designed to be applied to all people, regardless of an individual’s specific disability, the service received, the reason the service is being received, or the setting in which the service is received.
- **The Center for Disease Control’s (CDC) Healthy People 2010, Chapter 6 on Health and Equality for People with Disabilities.**² HP2010 Chapter 6 outlines solutions in the form of national goals and objectives for the United States addressing the unique needs of persons with disabilities. The recommendations, while placing focus upon persons with disabilities, are relevant to a broad spectrum of the U.S. and international population, including those persons being served in aging, employment, community living, behavioral health, drug treatment, and medical rehabilitation programs.
- **The Institute of Medicine’s (IOM) Crossing the Quality Chasm**⁷ provides specific guidelines for assessing and assuring quality health care in the United States. In this report, the IOM’s Committee on Quality of Health Care in America offers seven major recommendations as part of its overall “strategy and action plan for building a stronger health system over the coming decade.”⁷ Among the recommendations put forth by the committee is the “need for transparency. The health care system should make information available to patients and their families that allows them to make informed decisions when selecting a health plan, hospital, or clinical practice, or choosing among alternate treatments.”⁷ Furthermore, the report advocates the “incorporation of performance and outcome measures for improvement and accountability.”⁷

- **The Commission on Accreditation of Rehabilitation Facilities (CARF) International Board of Trustee’s Ends Policies** designed to promote and support CARF’s mission to enhance the lives of the person served. The **Ends Policies** serve to promote and support CARF’s mission by focusing on the following impact areas: (1) impact for the persons served by CARF-accredited programs and services, (2) impact from applying quality standards, and (3) impact for CARF accredited programs and services themselves. In addition, CARF Business Practices Standards relating to Information Management and Performance Improvement ask organizations to measure outcomes for persons served, including obtaining feedback from those served in any program seeking or maintaining accreditation. Because CARF must monitor progress toward these **Ends**, development of the uSPEQ questionnaire ensured that each of these areas is addressed by at least one item.

In addition, the construct underlying uSPEQ basically follows the classic Donabedian’s quality of care framework.⁴ More specifically, the survey items cover the domains of **access** (receipt of services), **process** (what happens during services), **outcomes** (results of services on the person served), and **structure** (organization’s capability to provide services). In this sense, part of the power of uSPEQ lies in its ability to benefit providers at a systemic level, making it an invaluable tool for all organizations.

The uSPEQ Consumer Experience Survey questionnaire consists of three tiers of items. They are Tier 1 (universal) items, Tier 2 (optional) items, and Tier 3 (custom) items:

Figure 1. uSPEQ Questionnaire with Three Tiers of Survey Items



The Tier 1 crosscutting items are universal for all populations in various settings of health and human services. These items are intended to measure the consumer's perceived service experience regarding the following five domains:

- Service responsiveness
- Informed choice
- Respect
- Participation outcomes
- Overall value

On the other hand, every provider is unique in many ways. Recognizing this, uSPEQ questionnaires can be customized to reflect the organization and program names relevant to each provider's data collection preferences. Providers can also choose to add Tier 2 optional items and/or Tier 3 custom items. Tier 2 items measure service experience important for one or more specific human service setting(s). Tier 3 items are custom items provided by the organization/provider to augment other aspects of service experience unique to the organization/provider, or to meet certain regulatory or funding requirements. These items are summarized in standard reports and are available for special reports as well.

Methods

Pilot Testing (uSPEQ v. 1.0)

Through a series of multi-stakeholder input forums and a consensus-oriented review loop, a pool of more than 80 items was generated for a pilot study of the uSPEQ questionnaire. uSPEQ v.1.0 was the product of the pilot testing in 2005, which included approximately 1,700 responses of consumers receiving services from 14 diverse CARF accredited organizations in 10 states. The distribution of respondents by provider service area or CARF customer service unit (CSU) is as follows:

Table 1. Distribution of Respondents by Provider Service Area

Service Area	# Respondents	Percentage
Aging Services (AS)	690	40.6%
Behavioral Health (BH)	243	14.3%
Employment and Community Services (ECS)	429	25.3%
Medical Rehabilitation (MED)	336	19.8%
Total	1,698	100.0%

The principal objectives of the field test were threefold:

- To refine the set of items by removing items that were not psychometrically fit for the instrument.
- To assess the validity and reliability of the uSPEQ questionnaire.
- To verify the feasibility of the survey process.

Statistical analyses were conducted in a planned progression. The analysis began with an examination of descriptive statistics that were produced for all demographic and questionnaire items (see **Appendix A**).

The pilot phase of uSPEQ questionnaire employed a five-point rating scale (i.e., 1 = **strongly disagree**, 2 = **disagree**, 3 = **neutral**, 4 = **agree**, and 5 = **strongly agree**). The majority (80 items) appeared on surveys for all pilot sites. These items are applicable across all components of human service continuum. Some items are applicable to most human service areas but not to all. Some items are primarily applicable to employment related programs or services, for example, **I am confident in my ability to use the skills I was trained in**. Still other items were specifically proposed by a pilot site and, therefore, were only applied to one provider. In addition to the questionnaire items, there were 18 demographic questions or questions related to the services received.

During the pilot testing phase, uSPEQ questionnaires were distributed on paper to the persons served at the pilot sites and, once completed, were returned to CARF for data entry and data analysis. The questionnaires were distributed by providing them at discharge, using them in interviews with non-direct service personnel, or mailing them directly to the persons served. The consumers completed the surveys and returned them to the organizations or directly to CARF International. The actual methods of survey administration varied slightly from site to site, depending on the specific circumstances of the pilot sites and characteristics of the service programs. Cognitive testing was conducted at several pilot sites, including focus groups with the survey respondents and extra questions accompanying the pilot questionnaire.

Pilot survey data were analyzed to assess the psychometric properties of the survey instrument. Correlational analysis, Rasch modeling, reliability analysis (Cronbach's alpha), and multiple regression, among other statistical procedures, were used to examine uSPEQ's validity and reliability, refine and reduce the instrument set, and ensure representation of the important constructs uSPEQ measures. Feedback from pilot sites' staff members and respondent focus groups (conducted on site right after the survey) also helped refine the instrument and data collection methodologies.

Rasch Modeling (Item Response Theory)

Rasch modeling is one of the psychometric models that helps with the evaluation of psychometric properties and creates measures and scales. Item Response Theory (IRT), upon which the Rasch model is based, provides an appropriate approach to handle ordinal item choice response. In addition, because uSPEQ is designed to measure service experience from persons served in various human service settings, it is essential that the instrument itself is invariant as well as sample independent. The Rasch model provides a comprehensive way to evaluate each item in the context of various service settings. It provides statistics for both items and persons to identify individual items that are not fit for the model.

Rasch analysis is a logistic item response model that constructs a line of measurement along which persons (e.g., respondents) and items (e.g., questionnaire items) are placed hierarchically using the same metric, an equal interval logit or log odds scale. On this same linear measurement scale, persons are ordered from less able to more able (in this case, less satisfied to more satisfied), and items are ordered from easy to hard (in this case, from being easy to endorse to being hard to endorse). The odds of a person endorsing a given item is modeled as a function of the person's overall level of ability (in this case, satisfaction with the service provided) and the difficulty of that item.¹⁵ Once the parameters of a Rasch model are estimated, they are then used to compute expected response pattern for each person on each item. Rasch modeling provides fit statistics that are essentially derived from a comparison of the expected patterns and the observed patterns.¹⁰ Two types of fit statistics, infit and outfit mean square statistics are usually reported by Rasch analysis programs to monitor the compatibility of the empirical data with the Rasch model. The infit is an information-weighted sum, sensitive to unexpected behavior affecting responses to items near the person's ability level. The outfit is based on the conventional sum of squared standardized residuals, sensitive to outliers. The planned target range of infit and outfit statistics is from 0.60 to 1.40.⁸

Item fit statistics are used to identify items that may not be contributing to a unitary scale (i.e., a violation of unidimensionality) or whose response depends on response to other items (i.e., a violation of local independence).¹⁵ In other words, a poor fit statistic for an item suggests that the item may not be related to the rest of the scale or may simply be statistically redundant with the information provided by other items.

With multiple rounds of Rasch modeling, 30 survey items on the original questionnaire were identified as misfit (i.e., infit statistics >1.4) for overall or by each service area, resulting in uSPEQ v.1.0 with 50 items. With reference to the results from exploratory factor analysis (SPSS, version 12.0), five factors were identified, namely, 1) Service responsiveness, 2) Informed choice, 3) Respect, 4) Participation, and 5) Overall value. Both the principal axis factor analysis and the principal component analysis suggested a 5-factor solution for the overall sample.

The following table gives the separation and reliability statistics for both persons and items for each factor from the Rasch modeling. Separation estimates the number of levels from 0 to infinity to which the distribution of persons or items can be reliably distinguished. Reliability, in addition, refers to the percentage of observed responses that are reproducible. Because both persons and items are placed on the same scale in Rasch modeling, reliability is estimated for items and for persons. The person measure reliability in Rasch is analogous to Cronbach's alpha in the Classical Test Theory (CTT). It estimates how well we can discriminate people based on their estimated visual ability. The item measure reliability indicates how well items can be discriminated from one another on the basis of their difficulty. Reliability ranges from 0.00 to 1.00. The closer the reliability is to 1.00, the less the variability of the measurement can be attributed to the measurement error. Usually, a person or item reliability over 0.80 is considered acceptable, indicating 20% item and person measure variability can be attributed to measurement error.

Table 2. Summary Reliability Statistics by Factor

Factor (# items)	Person		Item	
	Separation	Reliability	Separation	Reliability
Service responsiveness (10)	2.07	0.81	4.62	0.96
Informed choice (10)	2.17	0.82	5.02	0.96
Respect (9)	1.87	0.78	6.12	0.97
Participation (13)	2.32	0.84	6.32	0.98
Overall Value (8)	2.05	0.81	10.94	0.99
Whole instrument (50)	4.55	0.95	9.15	0.99

These 50 items were identified as Tier 1 items; i.e., they are applicable and were found fit for the 4 service areas that were field tested. In addition, another 29 items were identified to be Tier 2 items (one item was dropped because of its reverse wording). Unlike Tier 1 items, these items were found not fit for all 4 service areas. Of them, 25 items are applicable to more than 1 service area, while another 4 are applicable only for 1 of the 4 service areas.

Classical Test Theory Procedures

In addition, these 50 Tier 1 items were subjected to the Classical Test Theory (CTT) procedures; e.g., Cronbach's alpha (internal consistency measure for instrument reliability), convergent and discriminant/divergent validities, factor analysis for construct validity, and multiple regression analysis for predictive validity.

Cronbach's Alpha

The reliability of the 50 Tier 1 items established through different measures of internal consistency. One commonly used internal consistency measure is the Cronbach's alpha reliability measure. Cronbach's alpha for each subscale should be 0.80 or higher and for the entire questionnaire should be 0.90 or higher. uSPEQ v.1.0 has high Cronbach's alpha values, both for items within each domain and for all Tier 1 items altogether in the survey. For all uSPEQ 50 items, the Cronbach's alpha value is 0.977 (see Table 3). Domain-specific values range from 0.918 to 0.956. In addition, uSPEQ is a crosscutting instrument because it achieves high reliability when applied to different human service settings. The table below shows the versatility of uSPEQ; the Cronbach's alpha values for these settings range from 0.89 to 0.98.

Table 3. Cronbach's Alpha Values for uSPEQ Domains

Scale	# Items	AS	BH	ECS*	MED	Overall
Service Responsiveness	10	0.956	0.963	0.888	0.950	0.956
Informed Choice	10	0.953	0.957	0.921	0.951	0.948
Respect	9	0.946	0.920	0.906	0.927	0.931
Participation	13	0.951	0.933	0.892	0.917	0.933
Overall Value	8	0.914	0.915	0.902	0.938	0.918
All uSPEQ-50 Items	50	0.976	0.980	0.966	0.975	0.977

* Only 45 items of the uSPEQ-50 were examined for ECS crosscutting because of the small number of cases. The five items dropped were:

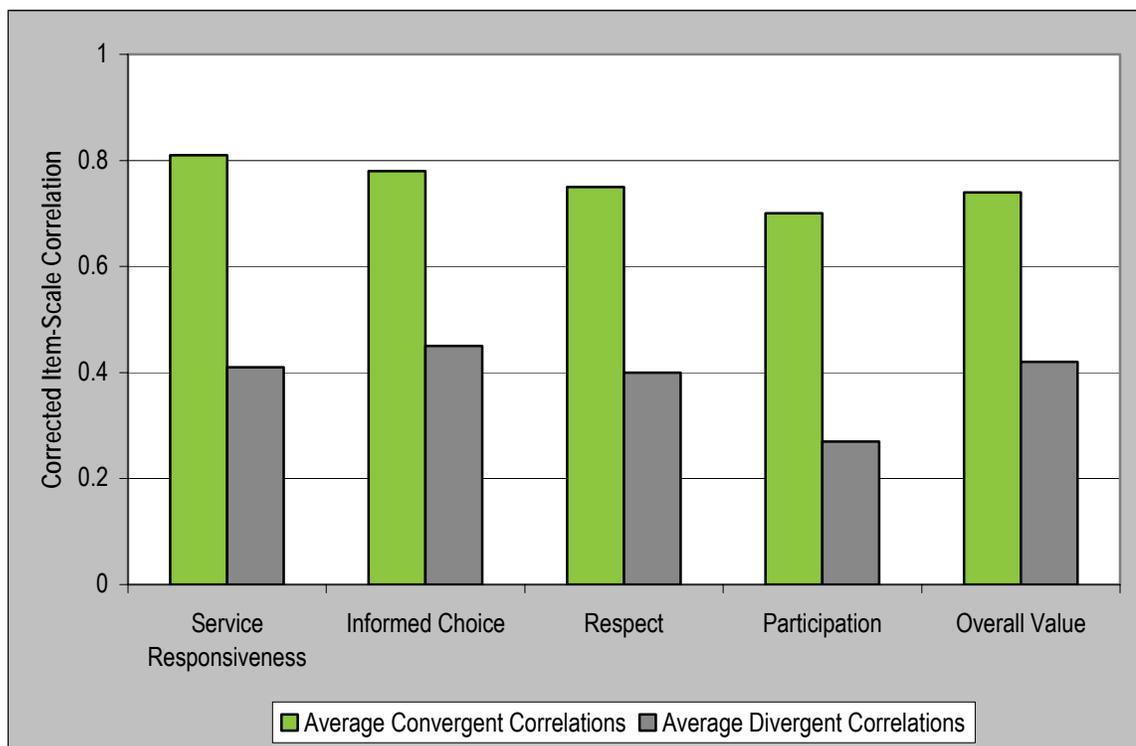
- I was served in a timely manner at [].
- I feel that people generally respect me even though I may have a disability.
- The services/care I received exceeded my expectations.
- As a result of the services I received, I am able to participate in leisure and recreational activities.
- I can choose to be as active as I want.

Convergent and Divergent Validity

In addition, intra-subscale (inter-item) correlation, corrected item-scale correlation, subscale-subscale and subscale-total correlation, and other internal consistency procedures were also run to further establish the reliability of the uSPEQ survey instrument.

Convergent validity examines how individual items are related to their own scale or domain. An individual item from a domain should be well correlated with the other items in the same domain. Conversely, discriminant or divergent validity examines how individual items in a domain are related to other domains. In general, an item should be more closely related to other items in the same domain or scale than to items in other domains or scales. The figure below provides a graphical contrast between convergent and divergent validities for each domain. Clearly, uSPEQ v.1.0 has demonstrated very good convergent validity and divergent validity.

Figure 2. Average Convergent and Divergent Validity



Predictive Validity

Predictive validity indicates the relation between overall satisfaction scores and other scores that theoretically should be linked to satisfaction. For example, when a consumer is satisfied with the quality of the services, it should logically follow that he/she would recommend the service to his/her friends or family or would return to the provider when similar needs emerge. Prevailing practice suggests that high predictive validity is observed when there is a high correlation between satisfaction and intent to recommend. In uSPEQ v.1.0, the remaining 7 items in Overall Value predicts 62% of the response variance in the item **I would recommend this program to a friend or family member** ($R^2 = 0.624$, adjusted $R^2 = 0.62$). The following table presents the F-statistics for the 7-item multiple regression model.

Table 4. Multiple Regression F-Statistics for 7-Item Model

ANOVA						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	378.423	7	54.060	230.080	0.000
	Residual	227.679	969	0.235		
	Total	606.102	976			

DV: E.1. I would recommend this program to a friend or family member.

IVs: All other items (7) in Overall Value.

Another multiple regression model, using 49 items in uSPEQ to regress on the intent to recommend service/care to a friend or family member, shows a slight improvement in R^2 from 0.624 to 0.655. Nonetheless, because the number of variables entered in this latter model is much larger than the previous 7-item model, the adjusted R^2 turned out to be almost identical (Adjusted $R^2 = 0.612$). The following table presents the F-statistics for the 49-item model.

Table 5. Multiple Regression F-Statistics for 49-Item Model

ANOVA						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	160.637	49	3.278	15.331	0.000
	Residual	84.681	396	0.214		
	Total	245.318	445			

DV: E.1. I would recommend this program to a friend or family member.

IVs: All other items (49) in the questionnaire.

The Pearson-r correlation between this item (**I would recommend this program to a friend or family**) and the overall score (calculated with this item excluded) is 0.508 (Spearman correlation: 0.543).

Survey Readability

Given that uSPEQ survey respondents may likely include persons with intellectual difficulty, it is important to assess the readability of the questionnaire. A readability test was conducted through a Microsoft product which showed that people with a grade 4 education should be able to read and understand the survey items on uSPEQ based on the following measurements:

- Average sentences per paragraph: 2.1
- Average words per sentence: 6.3
- Average characters per word: 4.0
- Passive sentences: 3%
- Flesch reading ease: 79.5
- Flesch-Kincaid grade level: 3.7

Independent Survey Sample

Strong empirical evidence on uSPEQ instrument reliability and validity also came from an independent survey sample with over 3,000 cases in Hennepin County, Minnesota. In 2007, the Hennepin County Human Services and Public Health Department designed a consumer feedback survey, incorporating 25 uSPEQ items and its five questionnaire domains. This department serves people over the whole spectrum of health and social services. Over a three-month period, 3,000 cases were collected for the survey (which is ongoing). Psychometric analyses were conducted on the survey instrument as well as on those 25 uSPEQ items separately. The results demonstrated strong reliability (Cronbach's alpha for internal consistency) for those 25 items as a whole and in their own domains or scales (all indexes were over 0.90).

uSPEQ Instrument Refinement (v.2.0)

In early 2008, uSPEQ underwent another round of data analyses in an effort to reduce the length of Tier 1 items and, at the same time, to maintain its psychometric properties. Both qualitative and quantitative analyses were conducted. Based on the result of a survey among the uSPEQ Steering Committee members as well as input from the Hennepin County Human Services and Public Health Department research group, several models were proposed to review the Tier 1 universal items. Using the framework developed from the qualitative analysis as a guide, Rasch analysis and CTT were utilized to:

- Further reduce the set of items by removing items that were not a good fit psychometrically for the instrument and items that were overly redundant with other items.
- Validate the psychometric properties of the refined uSPEQ questionnaire.

The survey refinement analyses were conducted on the version 1.0 customers. All items on the uSPEQ questionnaire employ a five-point rating scale (i.e., 1 = **strongly disagree**, 2 = **disagree**, 3 = **neutral**, 4 = **agree**, and 5 = **strongly agree**).

The resulting version 2.0 represents a triangulation between the qualitative and quantitative analyses and consists of 20 Tier 1 items. The Rasch rating scale model was used to identify misfitting and overly redundant items in order to reduce the total number of Tier 1 items and as a method to examine the Tier 1 models identified in the qualitative analysis. CTT was conducted on the resulting version 2.0 to validate the psychometric properties.

Survey Distribution Methods

The actual methods of survey administration were selected based on the specific circumstances of the site and characteristics of the service programs. The majority of subscribers administered the questionnaires while services were ongoing or at discharge. Completed questionnaires were typically returned to a central place in those organizations and then shipped in bulk to the uSPEQ office for data processing. For a few subscribers, the questionnaires were mailed to the persons served with self-addressed, stamped envelopes. One subscriber requested uSPEQ to administer the survey on its behalf. In this instance, uSPEQ mailed questionnaires directly to the persons served by this organization. Responses were mailed back to the uSPEQ office. Both paper- or web-based questionnaire options were available to the uSPEQ subscribers for version 1.0. Paper questionnaires are returned to the uSPEQ office for data entry and data analysis.

Survey Sample

By the time the version 1.0 database was closed for this round of analysis, the data sample consisted of 2,439 participants from 17 subscriber organizations. Participant demographic characteristics are presented in Appendix B. The distribution of respondents by provider service area or CARF customer service unit (CSU) is as follows:

Table 6. uSPEQ v. 1.0: Distribution of Respondents by Provider Service Area

Service Area	# Respondents	Percentage
Aging Services (AS)	169	6.9%
Behavioral Health (BH)	859	35.2%
Employment and Community Services (ECS)	1,226	50.3%
Medical Rehabilitation (MED)	185	7.6%
Total	2,439	100.0%

Factor Analysis

Factor analysis is a statistical procedure employed to uncover or confirm relationships among many variables or survey items. This allows numerous inter-correlated variables or survey items to be grouped under fewer dimensions, called factors. To confirm the five domains of uSPEQ items identified during the pilot study phase (as part of construct validity), all 50 Tier 1 items were subjected to factor analyses. To be consistent with the method used in the pilot data analysis, the principal component analysis Promax with Kaiser Normalization was performed. Exploratory analyses were used to examine if the items were loaded to the five factors or domains. The outcomes from the factor analyses clearly indicated a 5-factor solution, matching exactly with the 5 domains identified in the pilot phase.

Rasch Modeling

The Rasch rating scale model was used to help prune the pool of items so that the instrument is efficient and effective in measuring the constructs. Item reduction was an iterative process. With multiple rounds of Rasch modeling, 20 survey items on the uSPEQ v.1.0 questionnaire were identified as the best fit for overall and by each service area. These 20 survey items constitute the new version of the uSPEQ Consumer Experience Survey, version 2.0. Below are the summary reliability statistics, by person and by item, from the Rasch analysis on uSPEQ v.2.0.

Table 7. uSPEQ v.2.0: Summary Reliability Statistics by CSU

CSU	Person		Item	
	Separation	Reliability	Separation	Reliability
Aging Services (AS)	2.87	0.890	3.97	0.940
Behavioral Health (BH)	2.99	0.900	5.81	0.970
Employment and Community Services (ECS)	3.23	0.910	6.99	0.980
Medical Rehabilitation (MED)	2.78	0.890	4.52	0.950
Whole sample	3.07	0.900	9.99	0.990

Most of the 30 items that were excluded from the uSPEQ v.2.0 Tier 1 pool became Tier 2, optional items. A couple of items were eliminated completely from the instrument because they were so similar to other uSPEQ items. In addition, some new items were added to the pool of Tier 2 items. As a result, the pool of optional items has increased to 85 items.

Classical Test Theory Procedures

The psychometric properties of uSPEQ v.2.0 with 20 Tier 1 items were evaluated with some CTT procedures; e.g., Cronbach's alpha (internal consistency for scale reliability), convergent and divergent validities, and multiple regression analysis for predictive validity. Item response rates were analyzed for each and every questionnaire items as a method for identifying potential problematic items. The average item response rate was 97.4%, ranging from 96.5% to 98.5%.

Cronbach's Alpha

Cronbach's alpha is one of the most popular methods to measure the reliability of a survey instrument; e.g., in quantifying the reliability of a score to summarize the information of several items in questionnaires. It indicates the extent to which a set of survey items can be treated as measuring a single latent construct. While a reliability of 0.70 is considered adequate for a survey instrument, it is desirable for each subscale to be at 0.80 or higher, and for the entire questionnaire to be 0.90 or higher. uSPEQ v.1.0 with 50 Tier 1 items has high Cronbach's alpha values, both for items within each domain and for all Tier 1 items altogether in the survey. For all 20 uSPEQ items, the Cronbach's alpha value is 0.96 (see Table 8). Domain-specific values range from 0.81 to 0.92. Note that, typically, Cronbach's alpha values decrease with fewer items in a scale or subscale. In the next table, the alpha values as well as number of items are compared between the two versions.

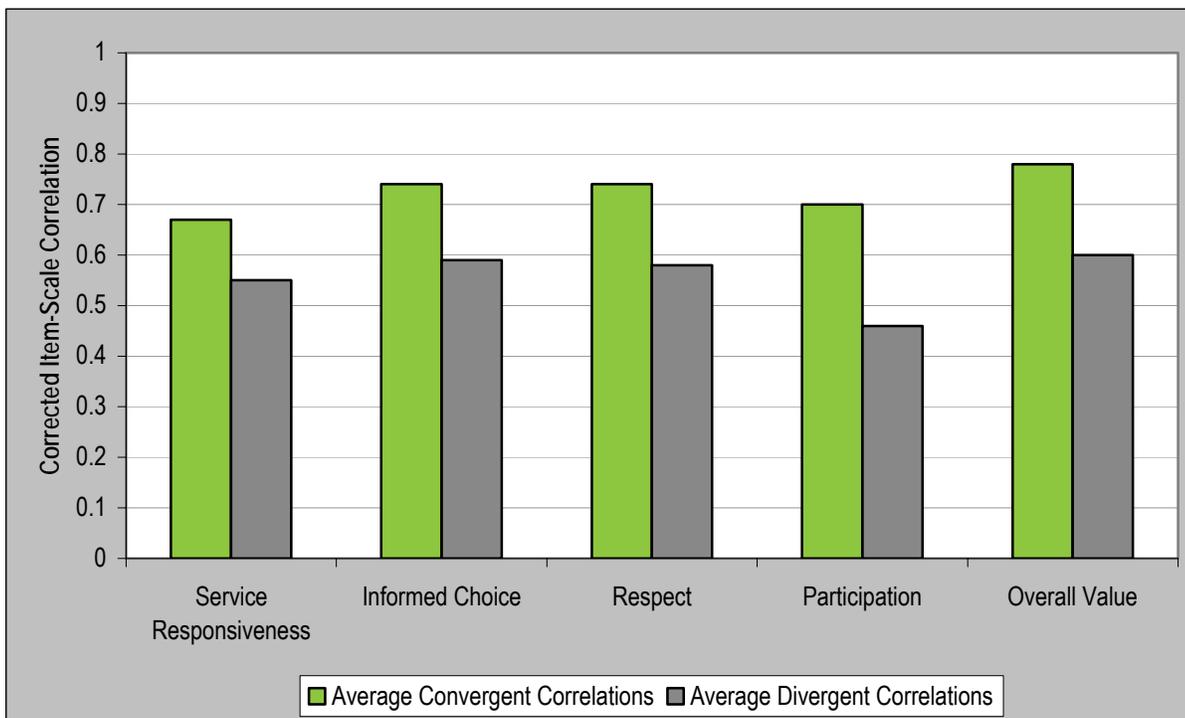
Table 8. uSPEQ v.2.0: Reliability Statistics (Cronbach's Alpha)

Scale/subscale	v.1.0: 50 Items		v.2.0: 20 Items	
	# Items	α	# Items	α
Service Responsiveness	10	0.940	3	0.810
Informed Choice	10	0.950	5	0.900
Respect	9	0.940	3	0.860
Participation	13	0.950	4	0.860
Overall Value	8	0.940	5	0.920
Entire Scale:	50	0.980	20	0.960

Convergent and Divergent Validity

The construct validity is considered the most valuable indicator of the validity of a survey instrument. There are two forms of the construct validity, namely, convergent validity and divergent validity. Similar to uSPEQ v.1.0, the 20 Tier 1 items in version 2.0 were evaluated for the construct validity of the new version. Specifically, individual items were examined to see if they were more highly related to the other items in the same domain (i.e., convergent validity) and, at the same time, not so highly related to items in other domains (divergent validity). The following bar graph contrasts convergent and divergent validities for each of the five domains. It is clear that all the convergent validities are higher than the divergent validity for uSPEQ v.2.0.

Figure 3. uSPEQ v.2.0: Average Convergent and Divergent Validity



Predictive Validity

As a form of criterion validity, predictive validity of a survey instrument is a measure of agreement between results obtained by the evaluated instrument and results obtained from more direct and objective measurements; for example, a re-purchasing behavior. Of the 20 Tier 1 items in uSPEQ v.2.0, item E.1. **I would recommend this program to a friend or family member** is perhaps the closest to a re-purchasing commitment or behavior. A high correlation (>0.5) between intent to recommend and overall satisfaction indicates good predictive validity. For uSPEQ v.2.0, the remaining 19 items predicts 62% of the response variance in item E1 ($R^2 = 0.627$, Adjusted $R^2 = 0.624$). The table below presents the F-statistics for the multiple regression model with 19 items.

Table 9. uSPEQ v.2.0: Multiple Regression F-Statistics for 19-Item Model

	Model	Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1128.392	19	59.389	183.915	0.001
	Residual	670.049	2075	0.323		
	Total	1798.441	2094			

DV: I would recommend this program to a friend or family member.

IVs: All other items (19) on the questionnaire.

Rating Scale

The choice of a rating scale for a survey instrument will shape the information to be collected from the survey. Accurate and reliable survey data depend on a combination of correctly written items and proper survey administration, as well as an appropriate rating scale. The criteria for a good rating scale include, 1) Easy to understand by respondents; 2) Discriminate well between respondents' perceptions; 3) Easy to interpret the survey results; and 4) Have minimal response bias.

There is no definitive conclusion on a good rating scale. To choose the most appropriate rating scale for a survey instrument depends on the goals of the survey project, and on what will be done with the survey results. Scales commonly range from 2 to 10 points (i.e., rating categories). A 7- to 10-point scale may seem to gather more discriminating information, but respondents may not be able to actually discriminate among the differences. On the other hand, 2- and 3-point scales offer few discriminate values. The most common are 4- and 5-point rating scales.

Another major contention in the literature has to do the middle point on the rating scale. A middle point is commonly labeled **Neutral**, **Neither agree nor disagree**, or **Neither satisfied nor dissatisfied**. It provides respondents with moderate opinions a way out. Without a middle point, it may contribute some form of random or systematic error to the distribution of responses. In contrast, other researchers believe that people are rarely neutral or without an opinion and, therefore, that neutral option is unnecessary. Explicitly offering a middle position may significantly increase the size of that category. In his classic book *The Art of Asking Questions (Studies in Public Opinion)*, Stanley Payne points out, "If the direction in which people are leaning on the issue is the type of information wanted, it is better not to suggest the middle ground.... If it is desired to sort out those with more definite convictions on the issue, then it is better to suggest the middle ground."¹⁷

uSPEQ v.1.0 employed a 5-point rating scale with a **Neutral** middle point. During the two years the survey was administered, quite a few subscribers expressed their concerns on the middle point. They experienced difficulty in interpreting the responses under that category, and it was difficult for the management to come up with action plans as a result. (Similar concerns were made from subscribers about another uSPEQ product, Employee Climate Survey.) In addition, data manipulations were made to compare a 5-point to a 4-point rating scale during the Rasch analysis for uSPEQ v.1.0. The conclusion from the data analysis was that no additional information was gained with a **Neutral** middle point for uSPEQ. Consequently, an administrative decision was made to change the uSPEQ rating scale to a 4-point scale; i.e., without a **Neutral** rating category, for version 2.0.

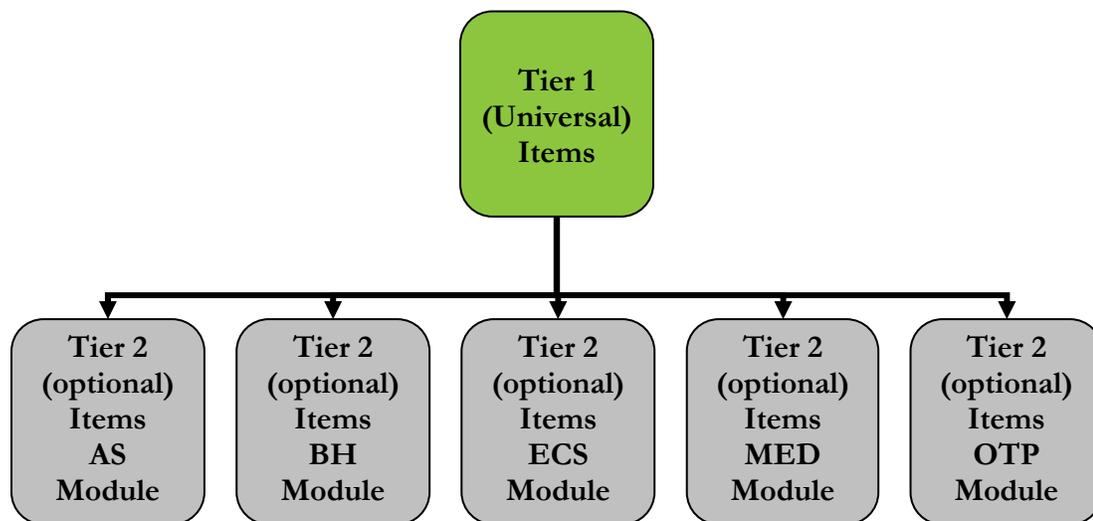
Conclusion

uSPEQ Consumer Experience Survey grew out of a decade of CARF’s extensive work on performance indicators. In developing uSPEQ, CARF was guided by feedback from providers; service payers; public agency representatives; researchers; and, most importantly, persons served by health and human service providers. In order to ensure that uSPEQ addresses areas of service experience relevant and important to the consumers, and from which items generated would accurately capture what they should be measuring, the development of uSPEQ underwent a long period of information collection. The questionnaire was developed based on the results of multiple workgroup meetings of various stakeholders as well as consumer focus groups and results from cognitive testing. Extensive reviews and crosswalks were conducted on the literature and active projects or efforts on and using performance indicators.

Since its public release in April 2006, uSPEQ Consumer Experience Survey has experienced a healthy growth, with subscribers most currently from 14 states and 2 countries in Europe. The uSPEQ v.1.0 with 50 Tier 1 items was the product of the field testing in 2005, which included 1,700 responses from the persons served in 13 voluntary organizations (pilot sites) from 10 states. More recently, the uSPEQ questionnaire was analyzed with another 2,500 cases, and as a result, further refined to 20 Tier 1 items in version 2.0.

The new version features a refined set of Tier 1 items (n=20) and an enriched set of Tier 2 optional items (n=85) presented in modules corresponding to service fields; e.g., aging services (AS), behavioral health (BH), employment and community services (ECS), medical rehabilitation (MED), and opioid treatment programs (OTP).

Figure 4. uSPEQ v.2.0: Tier 1/Tier 2 Modular Approach



The refined set of Tier 1 universal items will become the basis for future benchmarking. The optional Tier 2 items are grouped in modules by service field, which makes it easier for a subscriber organization to pick and choose from the optional item pool.

In general, the survey instrument development is in itself a continuous quality improvement process. With new survey data collected in the uSPEQ database, it will be possible to continue to refine the survey instrument over time. Efforts are underway to study and analyze new data in the future.

As a national and international pooled data set, uSPEQ's future will bring an opportunity for benchmarking. When the uSPEQ system has gathered sufficient data, there will be an opportunity to develop benchmarks, one of the key features for which providers have asked for help. Benchmarks can be used to compare the experiences of many people in many programs with the experiences of persons served by a specific program or in a particular community having specifically identified characteristics or across a broader field. The ability to understand how one group's experiences compare with those of others can help in understanding the needs, expectations, and challenges of people participating or residing in different settings. Providers want benchmarking to be in areas relevant to the persons they serve, their payers, and other stakeholders. They want to know about the average for all other providers or the range of acceptable values. Analyzing trends and differences between groups or over time requires a thorough understanding of the data plus sufficient data size to ensure meaningful sample sizes in various segments of the population(s). Research efforts in these important areas will continue using uSPEQ data as its use grows.

Appendix A: Descriptive Statistics on Pilot Study Participants

A total of 1,698 respondents rated the questions on the uSPEQ questionnaire and returned their survey to CARF during the pilot phase. Of them, 54.3% were females, 44.4% were males, and another 1.3% chose not to select on this item. These respondents were persons served from the 13 pilot organizations that voluntarily participated in the uSPEQ pilot study. The 13 pilot organizations were represented by 2 in aging services (AS), 3 in behavioral health and opioid treatment programs (BH), 5 in medical rehabilitation (MED), and 2 in employment and community services (ECS). Below is a table showing the number and percent of respondents by service area; i.e., CARF Customer Service Unit (CSU):

Table 10. Pilot Study: Respondents by Provider Service Area

Area	Count	Percent
AS	690	40.6
BH	243	14.3
ECS	429	25.3
MED	336	19.8
Total	1,698	100.0

Respondents by Age

The respondents' ages ranged from 17 to 100 years old, with a mean of 61.0 (SD = 24.6). See the following table for more details about the age of the respondents:

Table 11. Pilot Study: Respondents by Age

Age at time of completing uSPEQ	CARF CSU				Group Total
	AS	BH	ECS	MED	
Mean	83.8	41.4	33.0	65.0	61.0
Median	84.4	44.3	29.1	67.7	66.9
Minimum	56.6	17.4	17.3	19.1	17.3
Maximum	100.4	65.6	78.3	100.0	100.4
Std Deviation	6.6	11.1	13.6	16.6	24.6
Percentile 25	80.6	32.4	20.8	54.3	39.7
Percentile 75	88.3	49.3	43.4	77.7	83.2
Percentile 95	93.5	57.5	58.2	87.5	91.0

Respondents by Race and Ethnicity

Overall, most of the respondents (81.8%) were white, with 14.1% reporting Black or African American, 1.0% reporting Asian, 0.9% reporting American Indian or Alaska Native, 0.1% for First Nations/Aboriginal Canadians, 0.1% reporting other Pacific Islander, and 2.0% reporting other race. About 4.3% of the respondents reported they were of Hispanic or Latino ethnicity.

Table 12. Pilot Study: Respondents by Race

Race	CARF CSU									
	AS		BH		ECS		MED		Group Total	
	Count	Col %	Count	Col %						
White	646	96.6%	121	51.3%	328	76.8%	267	90.2%	1,362	81.8%
Black, African American	3	0.4%	101	42.8%	78	18.3%	53	15.9%	235	14.1%
American Indian or Alaska Native	2	0.3%	2	0.8%	9	2.1%	2	60.0%	15	0.9%
First Nations/Aboriginal Canadians	1	0.1%	1	0.4%					2	0.1%
Asian	6	9.0%	3	1.3%	3	0.7%	4	1.2%	16	1.0%
Other Pacific Islander	2	0.3%							2	0.1%
Other race	9	1.3%	8	3.4%	9	2.1%	7	2.1%	33	2.0%
Total:	669	100.0%	236	100.0%	427	100.0%	333	100.0%	1,665	100.0%

Respondents by Education

About half of the respondents had some college education (51.3%), 47.5% reported having up to a high school education, and 1.2% reported having no school education. A further detailed breakdown of education level is shown in the following table.

Table 13. Pilot Study: Respondents by Education Level

Level of Schooling Completed	Frequency	Percent	Valid Percent	Cumulative Percent
No schooling completed	20	1.2%	1.2%	1.2%
Nursery school to 4th grade	4	0.2%	0.2%	1.4%
5th or 6th grade	25	1.5%	1.5%	2.9%
7th or 8th grade	33	1.9%	2.0%	4.9%
9th grade	29	1.7%	1.7%	6.6%
10th grade	39	2.3%	2.4%	9.0%
11th grade	50	2.9%	3.0%	12.0%
12 grade (no diploma)	112	6.6%	6.8%	18.8%
High school diploma/GED	495	29.2%	29.9%	48.7%
Some college credit (less than one year or some trade school)	150	8.8%	9.0%	57.7%
One or more years of college (no degree or trade school certificate)	240	14.1%	14.5%	72.2%
Associate degree	54	3.2%	3.3%	75.5%
Bachelor's degree	220	13.0%	13.3%	88.8%
Master's degree	88	5.2%	5.3%	94.1%
Professional degree (MD, DDS, LLB, JD)	40	2.4%	2.4%	96.5%
Doctorate degree (Ph.D., Ed.D.)	22	1.3%	1.3%	97.8%
Other education	37	2.2%	2.2%	100.0%
Subtotal	1658	97.6%	100.0%	
Missing	40	2.4%		
Total	1698	100.0%		

Respondents by Completion/Help Methods

Close to two-thirds (63.3%) of the respondents completed the survey by themselves; 23.1% had someone else help them (reading the survey and/or writing answers) complete the surveys; and another 13.6% reporting someone else completed surveys on their behalf (i.e., surrogates).

Table 14. Pilot Study: Respondents by Completion/Help Methods

Person Completing Questionnaire	CARF CSU									
	AS		BH		ECS		MED		Group Total	
	Count	Col %	Count	Col %						
Myself–person receiving services (no one helped)	487	73.1%	201	87.8%	144	34.4%	201	63.0%	1,033	63.3%
Myself–someone helped me read and/or write my answers	66	9.9%	27	11.8%	175	41.9%	109	34.2%	377	23.1%
Someone else on behalf of the person served	113	17.0%	1	0.4%	99	23.7%	9	2.8%	222	13.6%
Total:	666	100.0%	229	100.0%	418	100.0%	319	100.0%	1,632	100.0%

Appendix B: Descriptive Statistics on Version 1.0 Participants

A total of 2,439 respondents in 17 organizations completed the uSPEQ v1.0 questionnaire. Of them, 40.4% were females, 53.6% were males, and another 6.0% chose not to answer this item. The 17 uSPEQ v.1.0 subscribers provided services in the following service areas: 2 in aging services (AS), 7 in behavioral health and opioid treatment programs (BH), 8 in employment and community services (ECS), and 3 in medical rehabilitation (MED). Three organizations had programs in more than one CSU (CARF Customer Service Unit) and were counted in both areas. Below is a table showing the number and percent of respondents by service area (CSU):

Table 15. Version 1.0: Respondents by Provider Service Area

Service Area	# Respondents	Percentage
Aging Services (AS)	169	6.9%
Behavioral Health (BH)	859	35.2%
Employment and Community Services (ECS)	1,226	50.3%
Medical Rehabilitation (MED)	185	7.6%
Total	2,439	100.0%

Respondents by Age

The respondents' age ranged from 12 to 101 years old, with a mean age of 44.4 years (SD = 17.5). See the following table for more details about the age of the respondents:

Table 16. Version 1.0: Respondents by Age

Age at time of completing uSPEQ	CARF CSU				Group Total
	AS	BH	ECS	MED	
Mean	77.6	41.4	40.7	56.7	44.4
Median	84.0	44.0	41.0	56.0	44.0
Minimum	24.0	12.0	15.0	19.0	12.0
Maximum	101.0	101.0	85.0	96.0	101.0
Std Deviation	18.4	13.7	14.1	19.7	17.5
Percentile 25	62.5	31.0	28.0	42.0	31.0
Percentile 75	90.0	52.0	52.0	73.0	54.0
Percentile 95	97.0	60.0	63.0	86.0	82.0

Respondents by Race and Ethnicity

Approximately two-thirds of the respondents (65.9%) were white, with 24.8% reporting black or African American, 1.6% reporting American Indian or Alaska Native, 1.5% reporting Asian, 5.9% reporting other race, and less than 1% First Nations/Aboriginal Canadians and or other Pacific Islander. About 8.3% of the respondents reported they were Hispanic or Latino ethnicity.

Table 17. Version 1.0: Respondents by Race

Race	CARF CSU									
	AS		BH		ECS		MED		Group Total	
	Count	Col %	Count	Col %	Count	Col %	Count	Col %	Count	Col %
White	148	96.7%	393	49.3%	850	72.6%	69	71.9%	1,460	65.9%
Black, African American	3	2.0%	277	34.8%	247	21.1%	22	22.9%	549	24.8%
American Indian or Alaska Native	1	0.7%	21	2.6%	13	1.1%	1	1.0%	36	1.6%
First Nations/Aboriginal Canadians	0	0.0%	1	0.1%	0	0.0%	0	0.0%	1	0.0%
Asian	1	0.7%	5	0.6%	25	2.1%	2	2.1%	33	1.5%
Native Hawaiian or other Pacific Islander	0	0.0%	5	0.6%	1	0.1%	0	0.0%	6	0.3%
Other race	0	0.0%	95	11.9%	34	2.9%	2	2.1%	131	5.9%
Total	153	100.0%	797	100.0%	1170	100.0%	96	100.0%	2,216	100.0%

Respondents by Education

About one-third of the respondents had a high school education (34.4%), 27.6% reported having less than a high school education, while one-third had at least some college or technical school training (34.4%). A further detailed breakdown of education level is shown in the following table.

Table 18. Version 1.0: Respondents by Education Level

Level of Schooling Completed	Frequency	Percent	Valid Percent	Cumulative Percent
8th grade or less	203	8.3%	9.2%	9.2%
Some High School but did not graduate	386	15.8%	17.5%	26.6%
Special Ed	22	0.9%	1.0%	27.6%
High School Diploma/GED	840	34.4%	38.0%	65.6%
Some college credit	343	14.1%	15.5%	81.1%
Technical/vocational school	17	0.7%	0.8%	81.9%
Associate degree	95	3.9%	4.3%	86.2%
Bachelor's degree	147	6.0%	6.6%	92.9%
Master's degree and above	82	3.4%	3.7%	96.6%
Other	76	3.1%	3.4%	100.0%
Subtotal	2,211	90.7%	100.0%	
Missing	228	9.3%		
Total	2,439	100.0%		

Respondents by Completion/Help Methods

Close to two-thirds (61.6%) of the respondents completed the survey by themselves; 25.9% had someone else help them (reading the survey and/or writing answers) complete the surveys; and another 12.5% reporting someone else completed surveys on their behalf (i.e., surrogates).

Table 19. Version 1.0: Study Respondents by Completion/Help Methods

Person Completing Questionnaire	CARF CSU									
	AS		BH		ECS		MED		Group Total	
	Count	Col %	Count	Col %	Count	Col %	Count	Col %	Count	Col %
Myself - no help	116	70.7%	678	82.0%	562	47.7%	89	50.0%	1,445	61.6%
Myself with help	34	20.7%	122	14.8%	392	33.3%	61	34.3%	609	25.9%
Someone else for me	14	8.5%	27	3.3%	224	19.0%	28	15.7%	293	12.5%
Total	164	100.0%	827	100.0%	1178	100.0%	178	100.0%	2,347	100.0%

References

1. Bond, T.G., and Fox, C.M. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Erlbaum, Mahqay, NJ.
2. Center for Disease Control and Prevention (2001). *Healthy people 2010*. Atlanta, GA: DHHS Publication.
3. Conrad, K.J., Wright, B.D., McKnight, P., McFall, M., Fontana, A., and Rosenheck, R. (2004). Comparing traditional and Rasch analyses of the Mississippi PTSD Scale: Revealing limitations of reversed-scored items. *Journal of Applied Measurement*, 5(1), 1-16.
4. Donabedian, A. (1980). *Exploring in quality assessment and monitoring*, Volumes 1-3, Ann Arbor, MI: Health Administration.
5. Kahler, C.W., Strong, D.R., and Read, J.P. (2005). Toward efficient and comprehensive measurement of the alcohol problems continuum in college students: The brief young adult alcohol consequences questionnaire. *Alcoholism: Clinical and Experimental Research*, 29(7), 1180-1189.
6. Green, K.G., and Frntom, CG. (2002). *Survey development and validation with the Rasch model*. Paper presented at the International Conference on Questionnaire Development, Evaluation, and Testing, Charleston, SC, November 14-17.
7. Institute of Medicine (2001). *Crossing the quality chasm*. Washington, DC: National Academy Press.
8. Linacre, J.M., and Wright, B.D. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8, 370.
9. Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danmarks Paedagogiske Institute. (Republished Chicago: The University of Chicago Press, 1980).
10. Smith, R.M. (1991). *IPARM: Item and Persons Analysis with the Rasch Model*. Chicago, IL: MESA Press.
11. Smith, E.V. (2001). Evidence for the reliability of measures and validity of measure interpretation: A Rasch measurement perspective. *Journal of Applied Measurement*, 2, 281-311.
12. Wilkerson, D.L., Shen, D., and Duhaime, M. (1998). *Performance indicators for rehabilitation programs*. Tucson, AZ: Commission on Accreditation of Rehabilitation Facilities.
13. World Health Organization (2001). *International Classification of Functioning, Disability and Health (ICF)*. Geneva: World Health Organization.
14. Wright, B.D. (1984). Despair and hope for educational measurement. *Contemporary Education Review*, 3(1), 281-288.
15. Wright, B.D., and Masters, G.N. (1982). *Rating Scale Analysis*. Chicago: MESA Press.
16. Wright, B.D., and Stone, M.H. (1979). *Best Test Design*. Chicago: MESA Press.
17. Payne, S. (1951). *The Art of Asking Questions (Studies in Public Opinion)*. Princeton: Princeton University Press.