

## APPENDIX V

Excerpted from Report to the Legislature on the Agency's Implementation of the Questionnaire for Situational Information (QSI) Assessment, submitted October 1, 2009  
Full studies are available at <http://apd.myflorida.com/qsi-wsc-training/>

### ***Validity and Reliability of the QSI***

Section 393.0661(1) (a), F.S., requires the Agency to use an assessment instrument that is reliable and valid.

APD has had five (5) studies conducted on the QSI to determine its reliability and validity in meeting the needs of APD customers. The Agency contracted with the Florida Center for Inclusive Communities (FCIC) at the University of South Florida to coordinate and conduct these studies of the QSI. Dr. Susan Havercamp served as the principal investigator. The FCIC subcontracted with the American Association on Intellectual and Developmental Disabilities (AAIDD) and the Human Services Research Institute (HSRI) for their assistance and support in this research. AAIDD is a professional membership organization, and HSRI is a consulting firm.

Final reports have revealed that overall the QSI has met reliability and validity standards; that is, the QSI generally measures what it is intended to measure and does so consistently across time and across the assessors administering the assessment. These standards measure test-retest reliability, inter-interviewer reliability, and concurrent validity. Additionally, the findings support that the three (3) subscales within the QSI which measure a customer's functional, physical, and behavioral status possess ample face and content validity; that is, they measure the types of things that are useful in planning supports. Further, the QSI uses many of the same criteria to assess its customers that are used by other similar validated instruments, such as the Supports Intensity Scale (SIS) and the Inventory for Client and Agency Planning (ICAP).

After initially establishing an instrument's validity and reliability, developers then begin a continuing process of alternatively designing enhancements to an instrument to improve validity and reliability and then testing the extent to which the enhancements do in fact improve validity and reliability. The Agency is committed to ongoing improvement of the QSI so it meets the needs of both the Agency and its customers. Accordingly, the Agency has conducted two (2) follow-up studies with the goal of enhancing the QSI.

The results are discussed below.

#### **Validity Studies**

**Validity** determines whether an assessment instrument measures what it was intended to measure. For a needs assessment instrument, a validity study is conducted to determine if the instrument accurately determines the needs of the individual in the areas that it

measures. For the QSI, this would be functional, behavioral, physical (health) status and the overall level of need. For example, one question to be answered might be “do the QSI’s questions regarding an individual’s medical circumstances accurately reflect his or her health needs”?

The QSI incorporates material from an assessment instrument previously developed by the Agency called the “Florida Status Tracking Survey” (FSTS), as well as new questions to make the assessment more comprehensive and updated. The Agency had conducted studies to assess the validity of the FSTS. These studies established acceptable construct validity (the FSTS measured the concepts it intended to measure) and concurrent validity (the FSTS accurately predicted the similar value contained in another instrument) for the FSTS.

Recent studies assessed content, face, construct, and concurrent validity of the QSI. A description of the studies and the results obtained thus far follows.

#### a. Content, Face and Construct Validity

Content validity is the systematic examination of the needs assessed so that items or questions selected for inclusion in an instrument represent what is intended to be measured. In the case of the QSI, the items should measure functional status, behavioral status, and physical status.

Face validity, as part of construct validity, concerns whether the assessment seems to contain the questions that pertain to the needs of people with disabilities. Construct validity is the extent to which the test or instrument measures a desired theoretical construct or trait. For the QSI, this construct would be the need for assistance or support.

The University of South Florida subcontracted with HSRI for a group of ten (10) content experts to analyze content, face, and construct validity. This group included both self-advocates and professionals in the field of developmental disabilities. The content experts were provided with a packet containing the QSI, the training manual, a brief description of the instrument’s history along with a letter describing the purpose of the study, and instructions to complete the review.

Regarding content validity, the investigators concluded that “the items in each scale are representative of the topic areas. Reviewers, however, sometimes observed that the scales could be expanded to cover additional ground within a topic area [functional, behavioral, or physical status]. On balance, however, the reviewers indicate that the items within each scale amply cover essential aspects within each targeted area.”

Regarding face validity, the investigators stated that the QSI’s three (3) scales (functional, behavioral, and physical status) appear to measure what is intended: the items in each scale do refer to the topic area targeted.

The investigators stated that the study offers some support for the construct validity of the three (3) scales. First, the scales have both face and content validity, a requisite for having construct validity. Second, reviewers stated that the QSI's scales compared favorably with those of similar widely-used needs assessment instruments like the Supports Intensity Scale (SIS) and the Inventory for Client and Agency Planning (ICAP).<sup>1</sup> The investigators indicated that "construct validity for any measure is developed over time as a preponderance of evidence builds to illustrate that the measure is aligned with a targeted hypothetical construct, such as 'the need for assistance or support.'" They recommend that additional studies be done, for instance, to assess how well the QSI's scales compare to other tools like the SIS and ICAP that measure similar constructs. (Note that the concurrent validity study performed by the principal investigator found correlations between the scales of the SIS and the QSI "within the moderate range of correlation indicating a substantial relationship." Section b., below, discusses this study.)

Among reviewers' recommendations were that the QSI assessment's language be more "people first" oriented, that the scaling and weighting of questions in determining scores be reviewed, and that the QSI include more questions regarding habilitative needs, given the QSI's strong emphasis on adaptive daily living skills.

The scaling and weighting of questions in determining scores have been reviewed through a factor analysis and repeat item analysis by the principal investigator and are discussed below. Other recommendations are currently under review by the Agency. For instance, the Agency could revise the questions in the Community Inclusion section to address habilitative needs better. These questions are not currently considered in determining a customer's level score but are useful in planning supports.

#### b. Concurrent Validity

Concurrent validity is a type of predictive validity and is often used to determine if two (2) similar needs assessments provide similar results in assessing needs. Concurrent validity is used to demonstrate where a test correlates well with a measure that has previously been validated. The two (2) measures may be for the same construct, or for different, but presumably related constructs. Our study examined the correlation of the QSI with the previously validated Supports Intensity Scale (SIS).

The SIS is a tool designed to measure the relative intensity of support that an individual with a developmental disability needs to participate fully in the community. The SIS consists of three (3) sections:

---

<sup>1</sup> The ICAP measures both adaptive and maladaptive behaviors and gathers additional information to determine the type and amount of special assistance that people with disabilities may need. The ICAP features two sections, one each measuring adaptive behavior and maladaptive behavior.

The SIS is a tool designed to measure the relative intensity of support that an individual with a developmental disability needs to participate fully in the community. The SIS consists of three (3) sections:

- A support needs scale,
- A section related to protection and advocacy, and
- A section assessing exceptional medical and behavioral support needs.

- A support needs scale,
- A section related to protection and advocacy, and
- A section assessing exceptional medical and behavioral support needs.

The SIS has been used as a needs assessment tool in several other states and provides a good comparison tool for validity. If the QSI correlates well with the SIS, this indicates that the instruments measure similar constructs or characteristics and would have similar applications in planning supports for people with developmental disabilities.

This concurrent validity study was conducted through an examination of the analysis of variance (ANOVA) between the QSI and the SIS for a sample of 100 individuals. The University of South Florida contracted with the American Association on Intellectual and Developmental Disabilities, who completed 100 valid SIS interviews by October 2008.

The principal investigator examined the Pearson product moment correlations<sup>2</sup> for the QSI functional scale, the QSI behavioral scale, the QSI physical scale, and the QSI total score with the corresponding scales and score of the SIS. This analysis compares how the change in one variable relates to the change in a corresponding variable, with a correlation of 1.0 indicating that given a change in one variable, the other variable changes by the same amount in the same direction. The investigator found that these correlations ranged from .59 to .66. Under widely accepted statistical standards, a correlation above .35 is desired. Since these correlations were all above the .35 threshold, they demonstrated concurrent validity and were within the moderate range of correlation indicating a substantial relationship.

**Table 1. Pearson product moment correlations between the SIS and QSI**

	SIS Home	SIS Community	SIS Learning	SIS Employment	SIS Health Safety	SIS Social	SIS Section 1 Total	SIS Medical	SIS Behavior
<b>QSI Functional</b>	0.74	0.48	0.50	0.54	0.65	0.52	<b>0.66</b>	0.60	0.02
<b>QSI Behavior</b>	0.13	0.25	0.07	0.14	0.16	0.32	0.21	0.08	<b>0.63</b>
<b>QSI Physical</b>	0.52	0.26	0.28	0.36	0.53	0.37	0.46	<b>0.59</b>	0.15
<b>QSI Total</b>	0.61	0.45	0.38	0.46	0.59	0.55	<b>0.59</b>	0.54	0.39

*Expected strong correlations appear in bold typeface and are shaded. Other correlations were not predicted to be strong and thus are not considered in determining concurrent validity.*

### **Reliability Studies**

<sup>2</sup> See Glossary on page 26 for definition.

For an assessment to be **reliable** it must first be found to be consistent in its measurement across time and across interviewers. Regarding the time element, the question to be answered is “are QSI assessment results the same when the QSI is administered to an individual at one point in time and then re-administered to the same individual at a later point in time, provided there has been no change in that individual’s situation”? In regard to the interviewer element, the question to be answered is “does the assessment obtain sufficiently similar results when administered to the same individual by different interviewers”?

Three (3) reliability studies have been completed regarding the QSI: test-retest reliability, inter-interviewer reliability, and item analysis.

#### a. Test-Retest Reliability

Test-retest reliability is a process used to assess the consistency of a measure from one time to another. In this test, QSI assessments were administered twice by the same QSI administrator for the same individual within a 2-3 week time period. The resulting scores of the two (2) QSI assessments were compared and analyzed. A high agreement between the scores from the two (2) assessments indicates strong test-retest reliability.

The Agency conducted 136 assessments statewide specifically for this study. Eleven (11) assessments were eliminated from the study as they had missing or unusable data. Initial data was sent to the principal investigator on August 22, 2008, and requested updated data and descriptive data was forwarded to the principal investigator on September 22, 2008.

The principal investigator examined test-retest reliability for two (2) groups: the total group of 125 valid assessments and a subsample of those who had not had major life changes during the interim between the two (2) administrations of the assessments, comprised by 111 persons. This subsample was examined since having a major life change could lead to legitimately different results in a second QSI administration.

The principal investigator reported that the test-retest reliability coefficients for both groups met or exceeded required thresholds (research standards) and were comparable to those reported for similar needs assessment instruments, including the Supports Intensity Scale (SIS), the Service Need Assessment Profile (SNAP), and the North Carolina Support Needs Assessment Profile (NC-SNAP).<sup>3</sup>

As shown in Table 2, scores were highly stable over the interval of several weeks for both the total group and the “no life changes” subsample. Pearson product moment correlation coefficients ranged from .86 to .94 for the entire sample and from .88 to .94 for the no life change subsample. Suggested reliability should generally be .80 or above for psychometric instruments (Anastasi and Urbina in Havercamp, 2009). As expected, test-retest correlations

---

<sup>3</sup> The SNAP is an Australian instrument designed to measure the support hours needed by individuals with disabilities living in the community. The NC-SNAP is intended to assess an individual’s intensity of need for services.

were greater for the subsample of individuals who did not experience a major life change during the interim between the two (2) administrations of the assessment.

**Table 2. Pearson product moment correlations between QSI time 1 and time 2**

	<b>Total Sample N=125</b>	<b>No Life Changes N=111</b>
<b>Functional</b>	.94	.94
<b>Behavioral</b>	.87	.90
<b>Physical</b>	.90	.90
<b>Level Estimate (Overall)</b>	.86	.88

b. Inter-Interviewer Reliability

Inter-interviewer reliability is used to assess the degree to which different raters (the QSI administrators, in the case of the QSI) give consistent ratings of the same individual using the QSI assessment.

QSI administrators completed a sample of fifty (50) assessments for use in inter-interviewer reliability studies by July 25, 2008, and the Agency provided data to the principal investigator for analysis by October 1, 2008. Results are shown in Table 3. The accepted industry standard for coefficients for inter-interviewer reliability developed by Cicchetti and Sparrow (1981) are 0-.39, Poor; .4-.59, Fair; .60-.74, Good; and .75-1.00, Excellent. The total score inter-interviewer reliability correlation was .74 (sum of scores), and the scales showed correlations at .87 for functional status, .48 for behavioral status, and .78 for physical status. The reliability correlation for the estimated level (overall score) was .45. To improve the behavioral and estimated level reliability correlations, the principal investigator suggested that APD conduct a factor analysis and repeat item analysis to further analyze how the QSI is scaled and weighted to determine the overall estimated level and specific level scores. Therefore, APD has contracted for these studies, the results of which are discussed below.

**Table 3: Pearson product moment correlations (Pearson's r) between QSI time 1 and time 2 (n=50)**

<b>Scale</b>	<b>Pearson's r</b>
Functional	.87
Behavioral	.48
Physical	.78
Total (Sum of scores)	.74
Estimated Level (Overall score)	.45

### c. Item Analysis

Item analysis is used to show how items relate to each other and the scores to which they contribute. Internal consistency considers the contribution of a particular question or item to the overall score. In other words, how does an instrument's validity improve given the addition of a particular question? Item agreement measures the extent to which various questions or items on an assessment agree; basically, whether some questions measure the same characteristics of an individual. Item discrimination determines whether a particular question contributes to the discrimination or determination of the overall score.

In September 2008, the principal investigator began a statistical analysis of the internal consistency, item analysis, and item discrimination of the QSI. For use in this process, Agency staff provided scores from a random sample of 500 assessments to the principal investigator on September 2, 2008.

The principal investigator examined the internal consistency of the QSI's three (3) scales. Internal consistency is a measure of the similarity of elements of the items on the scale. The report revealed that the QSI functional status scale had an acceptable internal consistency coefficient. However, the other two (2) scales (behavioral and physical) had internal consistency coefficients below the accepted standard. The thirty-six (36) items of the QSI which contribute to the estimated level had an internal consistency coefficient that approached the minimum level (.84, compared to a desired threshold of .85).

The principal investigator discussed two possible reasons for these findings. One is that the physical and behavioral subscales may combine two different types of items: items which measure support needs and items describing individual characteristics. By measuring a concept in different ways, more variation is introduced into the measurement. Another possible reason for these findings is that items in a single scale might be measuring not a single concept but more than one concept.<sup>4</sup> For example, the calculation of the score for the physical subscale includes some behavior-related items; possibly one or more of these items might not be related closely enough to physical health to merit inclusion in calculating the physical score.

The principal investigator suggested two strategies for improving internal consistency. One is to conduct an exploratory factor analysis. This type of study examines interrelationships among the items or questions in the QSI instrument. The results would highlight the concepts the QSI is measuring. For instance, the Agency may find that the QSI's measurement of physical or behavioral status is too broad and could be more narrowly defined. By measuring these concepts in more specific ways, the internal consistency would be improved. As will be discussed below, APD contracted for an exploratory factor analysis and found that more specific measurement of the constructs is possible and does improve internal consistency. A second strategy is to rewrite some questions to be more similar in their approach to measurement. This strategy would require more extensive readministration of the QSI and so is being considered for longer-term implementation.

---

<sup>4</sup> In fact, results of a subsequent study (a factor analysis) suggest that the questions comprising the physical subscale may be measuring as many as five factors, or narrower individual concepts, supporting this theory.

### **Ongoing Research to Improve the Instrument**

Assessment instruments generally undergo a continual process of refinement. There are a variety of ways to improve an instrument, from changes in the way that assessors are trained or administer an instrument to revisions in the scoring formula to rewording of questions or additions or deletions of questions in calculating scores.

From its base of overall good validity and reliability, to seek to enhance aspects of the QSI's inter-interviewer reliability and internal consistency, the Agency contracted for an exploratory factor analysis and a repeat item analysis as recommended by the principal investigator. These studies were completed June 30, 2009. The purpose was to summarize the interrelationships among scale items in a concise but accurate manner to better understand and measure the underlying construct (e.g., support needs). Through this analysis, the principal investigator posited that the QSI is comprised of not three (3) but nine (9) factors which she titled "Community Participation," "Self-care," "Behaviors," "Valued Social Roles," "Employment," "Physical Health," "Emergency Health Needs," "Seizure Needs," and "Sensory Problems." (Note that the principal investigator included as part of this analysis some of the QSI questions that were not included in the three (3) scored subscales: questions in the "Community Inclusion," "Fulfillment of Valued Adult Roles," and "Employment Information" sections.) By classifying questions into the nine (9) factors in this proposed factor solution, every question seemed to contribute to the internal consistency of the QSI scale.

The internal consistency coefficient for the QSI under this proposed factor solution is .92, above the desired threshold. The internal consistency coefficient for the individual factors ranged from .18-.92.<sup>5</sup>

Additionally, under this proposed factor solution, the patterns of factor intercorrelations are generally consistent with current trends in services and supports for people with intellectual and developmental disabilities. For example, according to the principal investigator, "We see strong relationships between community participation, employment, and attaining valued social roles. We also see relationships between self-care, employment, physical health, seizure needs, and sensory problems." However, she suggested that the Agency further examine a small number of the questions which did not appreciably contribute to the proposed subscales to which they had been assigned in the proposed factor solution.

Given these encouraging findings, the Agency intends to refine the QSI by contracting for further work on rescoring and rescaling the QSI. Among the questions this work would address is whether to distribute the QSI's questions differently among the three subscales, create a different set of subscales with the QSI's questions, exclude questions that do not appreciably contribute to a subscale from calculating subscale scores, or change the scoring system from its 1 to 5 range. APD expects that upon the completion of work developing a new scoring and scaling system, the existing assessments would be able to be used and would simply be

---

<sup>5</sup> As noted, the principal investigator included some items from the nonscored sections of the QSI. The .18 result was in regard what she termed the Employment factor which incorporates some of these items. There appear to be some data anomalies with one of the three items comprising this factor; when that question was deleted from the factor, the internal consistency coefficient rose to .85, at the desired threshold.

rescored using the newly-developed system. The Agency would then contract for additional studies to evaluate the validity and reliability of the revised QSI instrument.