

**Evaluating the Reliability and Validity of the
Questionnaire for Situational Information: Inter - Interviewer Reliability**

Final Report

Submitted By:

Susan M. Havercamp, PhD
Florida Center for Inclusive Communities, UCEDD
University of South Florida
13301 Bruce B. Downs Blvd. MHC 2113A
Tampa, FL 33612-3807
(813) 974-7076
SHavercamp@fmhi.usf.edu

With help from:

Steven J. Mason
Public Health doctoral student
University of South Florida

Prepared in collaboration with:

Agency for Persons with Disabilities
4030 Esplanade Way, Suite 380 S
Tallahassee, Florida 32399
(850) 488-1552

January 26, 2009

Introduction

The Agency for Persons with Disabilities (APD) is the Florida program tasked with serving the needs of Floridians with developmental disabilities. Prior to 2004, APD existed as the Developmental Disabilities Program within the Department of Children and Families. APD works in partnership with local communities and private providers to assist people who have developmental disabilities and their families. APD also provides assistance in identifying the needs of people with developmental disabilities for supports and services. An instrument, the Questionnaire for Situational Information (QSI), was developed to be used in the needs assessment process. The QSI collects information on need across multiple domains as well as demographics and situational information such as where the person lives. The intent of the QSI is to gather information to help plan supports. APD is also exploring the use of the QSI among other pieces of data to develop individual budgets for people receiving developmental disability services. Like Florida, states across the nation have developed needs assessment procedures to help allocate resources to better meet the needs of individuals and make costs more predictable.

As recommended by the Standards for Educational and Psychological Testing (Joint Committee on Standards for Educational and Psychological Testing, 1999), APD commissioned a series of studies to evaluate the psychometric properties of the QSI before using this new instrument to make important decisions about individual services and budgets. The series includes the following reports: item analyses, inter-interviewer reliability, test-retest reliability, content validity, and concurrent validity. This report presents the inter-interviewer reliability for the QSI. The QSI is a multidimensional instrument that is designed to measure how much support the person is anticipated to need over the next 12-month period, as expressed by an overall score that is generated from three sub-scores, or factors. The three factors are as follows: *Functional Status*: need for assistance during the normal course of a routine day, *Behavioral Intervention and Support Status*: the extent to which the person

needs support to manage challenging behavior, *Physical Status*: life situations and physical conditions that may pose a need for medical interventions or health care for the person.

This report summarizes the reliability of the QSI across different administrators, or inter-interviewer reliability. Inter-interviewer reliability is a measure of the similarity of results when the instrument is completed by different interviewers. If the scores from different interviewers do not agree, it may indicate a problem with the instrument itself or with the training provided to the interviewer. Since the QSI is completed by certified QSI administrators, it is important to establish that the findings would be similar across different administrators. To test this, the QSI was completed twice by two independent administrators. The QSI administrators used the same sources of information (records and informants). This report examines the inter-interviewer reliability by sub-scale, total raw score for the instrument, and the 1-5 level generated from the QSI scores.

Methods

Measure

The Questionnaire for Situational Information (QSI) was developed by Florida's Agency for Persons with Disabilities (APD), with the assistance of a private contractor, Human Systems and Outcomes, Inc. The assessment instrument was the combination of several instruments used and revised by APD over the course of the past 10 years. In every administration of the Questionnaire for Situational Information personal information is gathered by an observation of and a face-to-face interview with the individual with a developmental disability, the individual's guardian, and the individual's family (if the individual is a child or has given permission). In addition, interviews are conducted with the individual's caregivers, health care personnel, behavior analyst, counselors, and other persons as appropriate. Finally, individual records are reviewed including recent assessments and progress notes from medical records, school records, previous support plans, and relevant

information from other collateral sources, as appropriate. The QSI focuses on the support needs in 3 domains: functional status, behavioral status, and physical status. The intent, then, is that the individual's support plan would be based on information from the QSI, the person's preferences, and the extent to which certain personal outcomes and basic assurances are being met.

Participants

These analyses were based on a sample of 50 individuals who were assessed twice on the QSI by two different certified administrators of the QSI. The participants lived in 3 counties of Florida: Escambia (58%), Okaloosa (26%), and Santa Rosa (16%). The racial background of the sample was predominately white (68%) and black (30%) with 2% of the sample reported as "unknown race". Nationality was reported as USA for 100% of the participants. The sample was roughly equally comprised of males (48%) and females (52%). Ages ranged from 16.3 years to 68.7 years with an average age of 43.3 (SD= 12.1). The majority of the sample had primary diagnoses of mental retardation (92%) with the remainder having a diagnosis of Cerebral Palsy (8%). Secondary diagnoses were listed as Cerebral Palsy (6%), epilepsy (6%), Developmental Disability Public Law eligible (38%), and no secondary disability (50%). The majority of participants lived in supported living (92%) with the remaining 4 individuals living in family home (2%), foster placement (4%), or independent living (2%). None of the assessments used to estimate inter-interviewer reliability were included in any other psychometric study.

Procedure

Area 1 agreed to help with this study (Pensacola and surrounding counties). They identified individuals who had recent QSI assessments on file. Individuals who agreed to participate in the study were scheduled for a second assessment. The two QSI administrators used the same sources of information (informants and records) to make their ratings. In one case, the guardian was sick on the

day of the second assessment but a follow-up phone call was made to the guardian to get the information needed. The two QSI administrators were not permitted to discuss the scores with each other at any time during the process. The elapsed time between the 2 assessments ranged from 11 to 104 days with an average of 65.9 days and a standard deviation at 24.7.

Data Analyses

The results were analyzed using SPSS version 16.0. Raw subscale and total scores were computed and used for these analyses. Pearson's coefficients were computed for the subscale score, total scores, and Estimatr level between the two assessments for each participant. These data were compared with the accepted industry standard for coefficients for inter-rater reliability developed by Cicchetti and Sparrow (1981): 0-.4 Poor, .4-.59 Fair, .60-.74 Good, .75-1.00 excellent.

Results

Results are as shown in Table 1. The total score inter-interviewer reliability coefficient was .74, and subscales showed coefficients at .87 for functional, .48 for behavioral and .78 for physical. The reliability coefficient for the estimated level was .45.

Table 1:

Subscale	Pearson's r
Functional	.87
Behavioral	.48
Physical	.78
Total	.74
Estimated Level	.45

Comparison to Similar instruments

In addition to the standards proposed by Cichetti and Sparrow (1981), the inter-interviewer reliability coefficients can be compared to those reported for other published needs assessment

instruments. For purposes of the inter-interviewer reliability, four instruments were used for comparison. It is important to note that the term “inter-rater reliability” is sometimes used in the literature when discussing what should correctly be termed “interviewer reliability”. Technically, inter-rater would mean measuring the extent to which an informant (e.g. the person’s mother) produces different results than another informant (e.g. the person’s father). Inter-interviewer reliability refers to the extent to which the person asking the questions and administering the test produces results similar to other administrator, and this is the psychometric properties being evaluated in this study.

The first, and most widely recognized, instrument is the Supports Intensity Scale, which was developed by the Association for Intellectual and Developmental Disabilities, and has been adopted by many states and regional entities as a tool for support planning and budget allocation. (Thompson, et al., 2004) The recent inter-interviewer reliability study reported Pearson’s correlation coefficients ranging from .51-.92 (Thompson, Tassé, & McCaughlin, 2008).

The second instrument chosen as a comparison is the widely recognized Inventory for Client and Agency Planning (ICAP; Bruininks, Hill, Weatherman, & Woodcock, 1986). This instrument was developed in Minnesota in the 1980’s and continues to be used by a number of states, with various additions, for resource allocation purposes. According to a Buross Institute review of the ICAP by Wikoff (1989), inter-rater reliability for the ICAP ranged from .87-.92 on the various subscales.

A third instrument was recently developed in North Carolina Support Needs Assessment Profile, or the NC-SNAP (Hennike, 2006). The NC-SNAP is described by the state of North Carolina as a “needs assessment tool that, when administered properly, measures an individual’s level of intensity of need for developmental disabilities (DD) supports and services”. It was adopted for use for all individuals receiving services in North Carolina in 1999. The instrument was designed for several different disability groups, and includes supports across all domains, including medical needs and behavioral needs. The instrument uses information across a wide variety of needs to place individuals in

one of five service levels. Hennike, Myers, Realon, and Thompson (2006) reported on the psychometric properties of the NC-SNAP and the Revised NC-SNAP. They reported subscale inter-rater reliability coefficients for the revised NC-SNAP ranging from .84 to .88 for those in the intellectual disabilities group.

The fourth is an instrument developed in Australia, called the Service Need Assessment Profile. (Gould, 1988). This instrument is designed to measure support need for individuals with a variety of disabilities, and validity and reliability testing has consistently shown good results. Inter-interviewer reliability for the intellectual disability group ranged across subscales from .65-.93 (Guscia, 2005).

Conclusions

The Questionnaire for Situational Information (QSI) was developed to be used in the needs assessment process for the purpose of planning supports and possibly in budget allocation. This study examined the inter-interviewer reliability of the QSI.

The reliability of the QSI for the total score was .74, the subscale coefficients ranged from .48 to .87. The inter-interviewer reliability coefficient for the estimated Level was .45. These results fall in the Fair to Excellent range according to Cicchetti and Sparrow (1981) and are comparable to the four other widely used support need instruments.

In November 2008, shortly after this study commenced and before the preliminary inter-rater reliability report was written, APD undertook efforts to improve the reliability of the QSI. APD initiated a pilot project introducing a “Gold Standard” re-certification process. The new certification involves a two stage process; the first stage involves having newly certified interviewers rate an individual case that was previously rated by a QSI assessor and the gold standard committee. A group discussion format is used to talk about the gold standard process and findings for the case. Each area of the QSI is discussed and feedback sought on scoring rationale. The second stage is having the QSI assessor observed during a QSI assessment and

rated on a 1-4 likert scale that is based on QSI Certification Interview check list and scoring rubric. The checklist and rubric list interview competencies and graduated examples of QSI interview mastery levels. This process is expected to enhance inter-interviewer reliability of the QSI in the future.

Smith (1986) published an extensive review of 24 studies on the effects of rater training on the rater reliability. He draws the following conclusion,

Generally speaking, the more actively involved raters become in the training process, the greater the outcome. Providing raters with the opportunity to participate in a **group discussion** along with **practice** and **feedback exercises** produces better results than presenting the training material to them through a lecture. Practice and feedback exercises appear to be a necessary ingredient for increasing accuracy in ratings (Smith, 1986).

APD has recognized the importance of such feedback and evaluation processes during initial training and on an ongoing basis. In addition to the two-stage certification process described above, APD holds bi-weekly conference calls with QSI administrators and trainers. The purpose of these calls is to discuss and reach consensus on the administration and interpretation of the QSI. Questions and answers that arise during these calls are subsequently posted on a website for future reference.

In addition to improvements in the training (Anastasi & Urbina, 1997; Urbina, 2004), it may be beneficial to examine the wording of certain items, particularly on the Behavior subscale, which had the lowest correlation for inter – interviewer reliability at .48. Consideration should be given to changing the wording on items to make them more objective (Nunnally, 1967). Another consideration stems from the fact that QSI scores are based on the interview in addition to other

sources of information and documents. The presence and quality of that input is important in ensuring the reliability of the instrument (Anastasi & Urbina, 1997; Nunnally, 1967).

Finally, while the reliability coefficients for the subscale scores and the total score showed encouraging results, the correlation of .45 on the estimated level may indicate a need to re-examine the use of the estimated level. Because the inter-interviewer reliability of the estimated level was considerably lower than the reliability of the total score and of the subscale scores, APD may consider re-examining how the estimated level is computed or using the total score or a combination of subscale scores when making decisions about resource allocation.

In summary, the QSI shows promise as a measure of support needs, and the results of inter interviewer reliability were within acceptable ranges.

References

- Anastasi, A. & Urbina, S. (1997). *Psychological Testing* (Seventh Edition). Upper Saddle River, NJ: Prentice-Hall, Inc.
- Bruininks, R. H., Hill, B. K., Weatherman, R. F., & Woodcock R. W. (1986). *Inventory for Client and Agency Planning*. Itasca, IL: Riverside Publishing Company.
- Buntinx, W. (2008). The Dutch Version of the Supports Intensity Scale. In R.L. Schallock, JR. Tompson, and M.J. Tassé (Eds.). *AAIDD White Paper: Psychometric Properties of the Supports Intensity Scale*. Washington, DC: AAIDD.
- Cicchetti, D. V., Sparrow. (1981) Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*, 86 (2), 127.
- Gould, A. (1998). *The service need assessment profile* (Version 1.0). Sydney: ATG & Associates Pty Ltd.
- Guscia, . (2005). Reliability of the service need assessment profile (SNAP): A measure of support for people with disabilities. *Journal of Intellectual Developmental Disability*, 30(1), 24.
- Hennike, . (2006). Development and validation of a needs-assessment instrument for persons with developmental disabilities. *Journal of Developmental and Physical Disabilities*, 18(2), 201.
- Joint Committee on Standards for Educational and Psychological Testing (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Nunnally, J.C. (1967). Psychometric theory. In N. Garmezy, H.R. Harlow, L.V. Jones, and H.W. Stevenson (Series Eds.). *McGraw-Hill Series in Psychology*. New York: McGraw-Hill Book Company.
- Smith, D.E. (1986) Training program for performance appraisal: A review. *Academy of Management Review*, 11, 22-40.
- Thompson, J.R., Bryant, B.R., Campbell, E.M., Craig, E.M., Hughes, C.M., Rotholz, D.A., Schallock, R.L., Silverman, W.P., Tassé, M.J., Wehmeyer, M.L. (2004). *Supports intensity scale user manual*. Washington, DC: AAMR.
- Thompson, J.R., Tassé, M.J., McLaughlin, C.A. (2008). Interrater Reliability of the Supports Intensity Scale (SIS). *American Journal on Mental Retardation*, 113 (3), 231-237.
- Wikoff, R.L. (1989). Review of Inventory for client and agency planning. In J.C. Conoley and J.J. Kramer (Eds.), *The tenth mental measurements yearbook*. Rolling Meadows, IL: Riverside Publishing.
- Urbina, S. (2004). Essentials of psychological testing. In A.S. Kaufman and N.L. Kaufman (Series Eds.) *Essentials of Behavior Science Series*. Hoboken, NJ: John Wiley & Sons.