



## National Core Indicators

### NCI Adult Consumer Survey: Development and Psychometric Properties

*(Updated 9/13/12)*

#### Brief Summary

The NCI Adult Consumer Survey was created with input from a Project Advisory Committee with extensive experience in developing instruments and methods used to measure service system outcomes. A comprehensive literature review of outcome-based research and evaluation also informed the process. The tool has undergone rigorous testing both during initial piloting and after significant revisions. A summary of previous reliability test results is provided below<sup>1</sup>.

#### Inter-Rater Reliability

Inter-Rater Reliability is a measure of the level of agreement between two raters observing the same behavior under the same conditions. Different methods can be used to analyze and describe the consistency between ratings. Some are more precise than others and take into account that agreement may be solely based on chance. One of the more rigorous methods is to compute a statistic called Cohen's Kappa. By convention, a Kappa score of greater than 0.70 is considered acceptable inter-rater reliability. Testing done in five states yielded the following results:

- In 1997, a pilot test of the NCI tool was conducted with 30 individuals in Connecticut. Inter-rater reliability resulted in 93% agreement between the two raters.
- In November 1998, inter-rater reliability data were collected in Nebraska as part of the Phase I field test. The inter-rater reliability test (N=25) resulted in 93% agreement between the raters, and an average kappa score of 0.794.
- In April 1999, an inter-rater reliability test of Phase II items was conducted with 27 individuals in Minnesota. An analysis of inter-rater reliability found 92% agreement between raters.
- In 2008, the survey underwent some revisions, and a pilot test was conducted with 16 individuals in Massachusetts. Inter-rater reliability tests of this sample resulted in an average kappa statistic of 0.90 across pairs of raters.
- Between February 2010 and October 2010 a total of 20 inter-rater reliability tests and

---

<sup>1</sup> Source: Smith, G. & Ashbaugh, J. (2001). National Core Indicators Project: Phase II Consumer Survey Technical Report.



observations were conducted by HSRI in New Hampshire. Interviewers and shadows had Kappa agreement percentages of 0.88 or higher.

- A reliability study was done in the State of California during their first year of statewide implementation. It was done through a subcontractor (Bill Allen) to HSRI. See Appendix E of this report for results: [http://www.dds.ca.gov/QA/docs/NCI\\_SurveyReportFY2010-2011.pdf](http://www.dds.ca.gov/QA/docs/NCI_SurveyReportFY2010-2011.pdf)
- A second reliability study was done in California during their second round of Consumer Surveys.

### Long Version

**Instrument Development and Properties.** The NCI Consumer Survey was developed with extensive involvement of a Project Advisory Committee (PAC) composed of state developmental disabilities agency staff with expertise in data management, policy development and program implementation and advisors with expertise in conducting outcome research to create instruments and methodologies with validity in representing the goals of contemporary services. Instrument building was also aided by a substantial literature review of outcome-based research and evaluation. Draft survey items underwent substantial field-testing and revision. In 1997, a pilot test was conducted with 30 individuals in Connecticut. A Phase II reliability sample was selected with 15 consumers who were expected to have difficulty responding and 15 consumers who were expected to have little difficulty in responding to the questions. Inter-rater reliability resulted in 93% agreement between the two raters (Smith & Ashbaugh, 2001). In November 1998, both inter-rater and test-retest reliability data were collected in Nebraska as part of the Phase I field test. The inter-rater reliability test ( $N=25$ ) resulted in 93% agreement between the raters, and an average kappa score of 0.794. Test-retest reliability ( $N=27$ ) resulted in 80% agreement between the two administration times, and an average kappa score of 0.502. In April 1999, an inter-rater reliability test of Phase II items was conducted with 27 individuals in Minnesota. An analysis of inter-rater reliability found 92% agreement between raters regarding the ability of respondents to understand and respond reliably to the subjective questions posed (Smith & Ashbaugh, 2001). In 2008, the survey underwent some revisions, and a pilot test was conducted with 16 individuals in Massachusetts. Inter-rater reliability tests resulted in an average kappa statistic of 0.90 across pairs of raters. Between February 2010 and October 2010 a total of 20 inter-rater reliability tests and observations were conducted by HSRI in another participating NCI state. Interviewers and shadows had Kappa agreement percentages of 0.88 or higher. Feedback from each field-test also contributed to improved wording of questions and response codes. Field-testing also led to survey construction and training improvements, such as clarifying "not applicable" and "skips" more clearly, revising consistency check questions, and establishing more concrete rules for coding responses.



**Validity.** The survey was developed with extensive involvement of the Program Advisory Committee of state officials and other advisors to assure the NCI survey validly represented the established national goals of services for persons with ID/DD. Additionally, the draft survey was reviewed by a focus group of individuals with intellectual disabilities to pre-test the face validity of the questions. Focus group participants highlighted problematic questions, identified words that needed further definition, and suggested alternative ways of phrasing questions. These modifications were incorporated into the final survey.

To assure individual interview validity, interviewers are asked to give formal feedback on every interview conducted. At the end of Section I, there are two questions that ask the interviewer to make a judgment about the individual's comprehension of the questions and consistency of responses. This information provides project staff with information on survey questions that were confusing to an interviewee and help to further clarify questions and interviewer instructions each year. These items are also used during analyses to identify respondents the interviewer doubted had to have understood questions adequately to have responded reliably and validly.

Smith, G. & Ashbaugh, J. (2001). National Core Indicators Project: Phase II Consumer Survey Technical Report. Human Services Research Institute. Retrieval from: <http://www.hsri.org>.