

Interrater Reliability of the Supports Intensity Scale (SIS)

James R. Thompson

Illinois State University

Marc J. Tassé

University of South Florida

Colleen A. McLaughlin

University of Medicine and Dentistry of New Jersey, Robert Wood Johnson Medical School

Abstract

The interrater reliability of the Supports Intensity Scale (SIS) was investigated under the condition that interviewers had to have been trained and/or experienced in its administration and scoring. Both corrected and noncorrected Pearson's product-moment coefficients were generated to assess interinterviewer, interrespondent, and mixed interrater reliability. The correlation coefficients for the SIS Support Needs Index Score and SIS subscale scores were considerably higher than coefficients reported in the *SIS Users Manual* that were derived from a similar study conducted with untrained, inexperienced interviewers.

The Supports Intensity Scale—SIS (Thompson et al., 2004a) was developed “in response to changes in how society views and relates to people with disabilities” (Thompson et al., 2004b, p. 1). One of the trends influencing the development of the SIS was the emergence of ecological person–environment interaction models for understanding disability (see Luckasson et al., 2002; World Health Organization, 2001). By focusing on behaviors and contextual factors that impact an individual's functioning in community settings, proponents of these models suggest that disability is best understood as a mismatch between environmental demands and what an individual is able to do. Although everyone experiences some degree of mismatch between their performance and environmental demands at one time or another, advocates of ecological models contend that there is a certain threshold (i.e., extent of pervasiveness) where the mismatch precipitates the need for individualized supports that differ both quantitatively and qualitatively from the types of supports

needed by the general population. The SIS was developed to provide an objective and standardized instrument to measure the extraordinary support needs of a person with intellectual disability relative to others. It has become a popular tool among governmental agencies and provider organizations throughout the world, but especially in the United States and Canada (American Association on Intellectual and Developmental Disabilities, 2007).

Any measurement instrument that is used to inform decisions that impact the lives of individuals must be carefully examined and continually improved. Fundamental to a scale's validity is its reliability. Indeed, reliability is a prerequisite, a necessary but insufficient, condition for validity. Cicchetti and Sparrow (1981) provided the following guidelines for evaluating reliability coefficients for adaptive behavior scales: less than .40 is poor, .40 to .59 is fair, .60 to .74 is good, .75 or greater is excellent.

When evaluating reliability coefficients (r), it

is useful to examine both noncorrected and corrected coefficients (Joint Committee on Standards for Educational and Psychological Testing, 1999). When sampling procedures produce instances of restricted variance on certain variables, as was the case with the SIS, corrected r s can be estimated to correct for attenuation effects. In the *SIS Users Manual*, Thompson et al. (2004b) presented interrater reliability results for both Pearson's product-moment correlation and the corrected r .

The *SIS Users Manual* includes information on five indices of reliability, all of which are very strong, with the exception of the interrater reliability findings. Thompson et al. (2004b) concluded that SIS interrater reliability findings were "mixed" and noted that the interrater study "was a very rigorous test" (p. 103) because different interviewers had interviewed different respondents. Clay-Adkins (2004) reported that interviewers in the original study had not received any training on how to administer the scale; they had only informal instructions and were provided with concise written directions. At the time of data collection, there were not any training sessions available nor were there any background materials (e.g., *SIS Users Manual*) to which interviewers could refer. Additionally, because the SIS was not yet published, none of Clay-Adkins' interviewers had any prior experience with the instrument.

Our primary purpose in the present study was to investigate the interrater reliability of the SIS when completed by interviewers who had completed training on how to administer and score the instrument and/or had become familiar with administration and scoring guidelines through extensive use of this scale with multiple individuals. Because completing the SIS requires the involvement of interviewers (who score the scale) as well as respondents (who provide the information), a major goal for us was to separate the error variance attributable to different interviewers and different respondents.

Method

Participants

Participants in this study included interviewers, respondents, and people with intellectual disability. Three interviews were completed for each of the 40 cases, for a grand total of 120 interviews; 111 (93%) were completed by interviewers who attended a SIS training workshop conducted by one of the authors of this study. The 9 other in-

terviews (8%) were completed by individuals employed in organizations that were in the process of studying, piloting, or adopting the SIS. Organizational representatives reported that these interviewers had administered and scored the SIS on multiple individuals and were familiar with SIS administration and scoring guidelines; however, it is not known how these interviewers initially learned how to administer and score the scale. They may have attended some type of SIS training workshop, or they may have only studied information presented in the *SIS Users Manual*.

The participation of interviewers was completely voluntary. They were asked to collect data on at least one case, but they were given the option of completing work on more than one case. There were two interviewers assigned per individual assessed. In total, 51 individuals served as interviewers: 45 were from four states within the United States, 2 were from Australia, and 4 were from Canada. One interviewer collected information on 7 cases, 2 on 5 cases, 4 on 3 cases, and 7 on 2 cases; 37 interviewers collected data on 1 case. Demographic information collected on interviewers included the following: gender (80% female, 20% male), highest educational degree received (76% bachelor, 22% master, 2% doctorate), ethnicity (73% Euro American, 27% other), job title (26% case manager, 22% program coordinator/manager/director, 17% psychologist/social worker, 36% other), years of work experience (30% less than 3 years, 31% from 3 to 10 years, 39% more than 10 years).

Interviewers were asked to report the extent to which they knew the individual with the disability who was being evaluated prior to conducting their interview. As mentioned before, each case had two interviewers. There was considerable variation among pairs of interviewers with regard to prior familiarity with their subjects. Although at least one of the interviewers reported knowing a subject very well in 17 cases, in no cases did both interviewers report knowing a subject very well. In 9 cases both interviewers indicated that they had no significant knowledge about the person with the disability prior to completing the SIS. Guidelines provided in the *SIS Users Manual* do not require that an interviewer know the individual who is being assessed. However, it is essential that respondents have good knowledge/familiarity of the individual being assessed (Thompson et al., 2004b).

Respondents

Although we knew the names of the interviewers because they had volunteered to participate, we did not know the names of the respondents. Interviewers were told to identify 2 respondents for participation, one of whom would be interviewed separately by both interviewers and one who would only be interviewed by one interviewer. Although respondents' names reported to the researchers were pseudonyms, the interviewers were instructed to collect demographic information on each respondent, and this information was checked across data forms to assure that respondents were accurately accounted for in data collection. In total, 80 respondents participated. Demographic characteristics of these respondents were gender (79% female, 21% male), relationship to the individual with a disability being evaluated (53% direct support staff, 8% parent, 9% work/job supervisor, 31% other), ethnicity (59% Euro American, 24% African American, 18% other), length of time knowing the person with a disability (15% one year or less, 29% from 13 months to 3 years, 31% more than 3 years up to 6 years, and 25% more than 6 years). All respondents reported that they knew the person being evaluated well.

People With Disabilities

The support needs of 40 individuals with an intellectual disability were evaluated using the SIS. The SIS Support Needs Index Scores generated from these evaluations ranged from 63 to 137, with a mean and *SD* of 94.63 and 15.44, respectively. The demographic characteristics are shown in Table 1.

Procedure

Data-collection procedures. Pairs of interviewers were provided with packets containing specific directions on how to collect data, SIS scoring sheets, demographic information forms, and return envelopes. The instructions to the interviewer pairs were:

1. Identify a person with a disability on whom to complete the SIS. It is critical that all three of the SIS interviews are completed on the same person.
2. Write a fictitious name for the person with the disability on each SIS interview form, but be consistent with the fictitious name.
3. Identify respondents for SIS interviews (i.e., fill out the SIS form). The people who are interviewed are the respondents. Respondents must be individuals who are very fa-

Table 1. Characteristics of People Assessed

Variable	%
Gender	
Male	58
Female	43
IQ range	
51-69	33
36-50	18
20-35	13
<20	18
Missing data	20
Adaptive behavior deficit	
Mild	25
Moderate	15
Severe	13
Profound	25
Missing data	23
Other disabilities in addition to ID ^a	
Speech/language	43
Physical disability ^b	40
Psychiatric disorder	38
Chronic health condition	23
Physical disability ^c	20
Other	40
Ethnicity	
Euro American	53
African American	15
Other or unknown	33
Primary residence	
Paid staff	
6 or fewer	53
7 to 15	18
16 or more	5
Parents	13
On own or apt. with supports	13
Primary employment	
Sheltered	33
Competitive or supported	13
Unpaid volunteer	13
Unemployed	20
Other employment	23

^aIntellectual disability. ^bSuch as motor limitations. ^cSuch as arm/hand limitations.

SIS interrater reliability

J. R. Thompson, M. J. Tassé, and C. A. McLaughlin

- miliar with the consumer (i.e., person with a disability). Write fictitious names for respondents on all forms, but be consistent with these names.
4. Record demographic information on the interviewers, respondents, and person with disability on corresponding forms.
 5. Decide which interviewer is going to interview two different respondents and complete two SIS forms. Then, assign the other interviewer to interview one of the same respondents that the first interviewer interviewed. It is essential that one respondent agree to be interviewed twice regarding the same consumer.
 6. Conduct all SIS interviews independently. Do not discuss ratings with each other. When scheduling SIS interviews, please allow at least 1 day between interviews.
 7. Upon completion of any form/phase of data collection (i.e., SIS forms, demographic forms), return the completed materials to the researchers.

There were no follow-up efforts with potential interview pairs to collect information, and no records were kept with regard to how many interviewer pairs were invited to participate because open invitations were offered at the conclusion of training workshops. However, it is obvious that multiple pairs of interviewers did not elect to participate because more packets were distributed than were returned. Eight partial packets were received that included information from only one interviewer (i.e., the other interviewer did not follow through). Data from these packets were not included in the analysis.

Data-analysis procedure. We analyzed data from completed interviewer packets using the Statistical Package for the Social Sciences (SPSS 11 for Mac OSX). Pearson’s product-moment coefficients (*r*) and corrected Pearson’s product-moment coefficients (corrected *r*) were generated to assess

interinterviewer reliability (pairs of SIS scores generated from different interviewers who interviewed the same respondent), interrespondent reliability (pairs of SIS scores generated from the same interviewer who interviewed different respondents on two different occasions), and mixed interrater reliability (pairs of SIS scores generated from the different interviewers who interviewed different respondents). We compared reliability coefficients generated from these three data-collection methods in order to provide insight into variability that is attributable to having different interviewers and different respondents.

Results

Table 2 shows the results of interinterviewer, interrespondent, and mixed interrater reliability tests for the six SIS subscales as well as the SIS Support Needs Index Score (i.e., the composite score) using Pearson’s *r* and the corrected *r*. In addition, Table 2 displays the mixed interrater reliability scores from the study published in the *SIS Users Manual* (Thompson et al., 2004b) based on data collected by Clay-Adkins (2004).

Correlation coefficients generated from data in the current study are markedly higher than those reported in the *SIS Users Manual* (Thompson et al., 2004b). Based on Cicchetti and Sparrow’s (1981) guidelines, across all three conditions, the SIS Support Needs Index Score was in the excellent range using both corrected and noncorrected *rs*. Only three subscales had noncor-

Table 2. Reliability Coefficients (Pearson’s *r*) and Corrected Reliability Coefficients (Corr. *r*) for Interinterviewer, Interrespondent, and Mixed Interrater Reliability

Scale	Interrespondent ^a		Interinterviewer ^b		Mixed interrater ^c		Mixed interrater ^d	
	<i>r</i>	Corr. <i>r</i>	<i>r</i>	Corr. <i>r</i>	<i>r</i>	Corr. <i>r</i>	<i>r</i>	Corr. <i>r</i>
Home Living	.83	.73	.89	.80	.86	.76	.79	.90
Community Living	.85	.91	.85	.89	.83	.90	.56	.68
Life Long Learning	.60	.75	.73	.88	.51	.66	.35	.55
Employment	.74	.93	.54	.77	.47	.69	.36	.55
Health and Safety	.84	.91	.92	.96	.81	.90	.58	.72
Social	.65	.87	.51	.74	.70	.90	.36	.60
SIS Support Needs Index	.87	.87	.90	.88	.85	.83	.54	.59

Note. The *p* level was less than .01 (2-tailed) for all correlations.

^aSame interviewer with different respondents. ^bDifferent interviewers with same respondent. ^cDifferent interviewers with different respondents. ^d2004 Supports Intensity Scale (SIS) *Users Manual*.

SIS interrater reliability

J. R. Thompson, M. J. Tassé, and C. A. McLaughlin

rected r s in the fair range, and none were in the poor range.

Discussion

The findings from this investigation clearly suggest the value of trained interviewers. Although a definitive cause and effect knowledge claim cannot be made in regard to the relationship between training and interrater reliability, the reliability coefficients from the current study are far superior to those reported by Clay-Adkins (2004) and by Thompson et al. (2004b) in the *SIS Users Manual*.

Because both adaptive behavior scales and the SIS involve interviewing others to collect information about everyday activities, it is reasonable to assume that SIS interviewer qualifications would be the same as those for professionals who complete adaptive behavior scales. However, this assumption is probably wrong. Completing adaptive behavior scales largely involves making judgments about whether a person does or does not typically perform specific observable skills and tasks. In contrast, the SIS requires judgments to be made about the frequency, duration, and type of supports that an individual needs to successfully participate in life activities that involve multiple skills and multiple domains of personal competence. For example, under Part F (Social Activities) of the SIS, Item 4 is “making and keeping friends.” In describing this item, Eliason et al. (2005) indicated that it

includes supports to: (a) communicate with peers; (b) learn social skills related to friendships; (c) arrange meetings with others and joining activities with friends; (d) locate programs/activities that lead to friendships; and (e) use the phone and other communications to maintain friendships. (p. 13)

Determining the frequency, time, and type of support needed to participate in such a complex activity would certainly seem to require an interviewer to engage in more probing and clarifying than would determining whether an individual performs certain everyday skills in their repertoire of behavior. Therefore, the SIS may require more training than is needed for administration of adaptive behavior scales to ensure reliable administration because it is an instrument and construct that is quite different from the usual assessment methodology that human service professionals are accustomed to. Without training, a significant number of SIS interviewers may be inconsistent

in interpreting the life activities and support needs measures included in the scale.

The only reference to interviewer qualifications provided in the *SIS Users Manual* (Thompson et al., 2004b) is that the “SIS should be administered by a professional who has completed at least a 4-year degree program and is working in the field of human services (e.g., case manager, psychologist, social worker)” (p. 24). Additional language suggests that several years of direct experience in the field is desirable. Based on this study, the authors of the SIS may want to reconsider these requirements and specify that interviewers must complete training before administering the scale. Based on the results from the participants in this study, a one-day workshop may be sufficient.

These findings also have implications for decision-making by policymakers in the area of developmental disabilities. Although the SIS was not designed to be a tool for informing funding decisions, an increasing number of states have included SIS scores in establishing reimbursement rates and individualized resource allocations (Smith & Fortune, 2006). Determining fair and equitable models for distributing resources is an important and complex task (see Stancliffe & Lakin, 2005) that is well beyond the scope of this investigation. However, if policymakers have decided that measuring the intensity of a person’s support needs provides information useful in establishing reimbursement formulas or determining individualized resource allocations, they can be confident that the SIS is a reliable measurement tool.

As with any study in which researchers investigate the reliability of a measurement tool, it is important to remember that reliability does not assure validity. Although reliability is typically considered a prerequisite for validity, Greenspan (2006) argued that in some instances the quest for reliability might actually undermine efforts to achieve validity. According to Greenspan, the best way to attain validity for a scale “is to base it on the construct of interest, even if it may be a little hard, perhaps because of its fuzziness, to measure it reliably” (p. 180). Although the *SIS Users Manual* (Thompson et al., 2004b) provided evidence supporting the content, criterion-related, and construct validity of the scale, Harries, Guscia, Kirby, Nettelbeck, and Taplin’s (2005) investigation suggested that the SIS may not be measuring a construct different than what is already assessed by

SIS interrater reliability

J. R. Thompson, M. J. Tassé, and C. A. McLaughlin

two popular adaptive behavior scales. Additional investigations focusing on the validity of the SIS as well as alternative approaches to measuring support needs are necessary.

Limitations

The findings from this investigation should be considered in light of at least two limitations. First, interviewers only interviewed one respondent when completing the SIS. This was done in order to isolate error variance due to having different respondents. The *SIS Users Manual* clearly calls for interviewers to collect information from at least 2 respondents when completing the SIS and incorporate the responses of those interviewed into a single assessment form. It is possible that if interviewers in this study had interviewed multiple respondents instead of a single respondent, the reliability coefficients would have been different. Because our results showed such a negligible difference between error variance attributable to different respondents and error variance attributable to different interviewers, future reliability studies of the SIS should contain multiple respondents.

A second limitation is that the interviewers were volunteers as opposed to a random sample of interviewers. These interviewers were willing to do extra work without any additional compensation because they wanted to contribute to the knowledge base of their field. It is reasonable to assume that the volunteer interviewers were individuals who, because of their high level of commitment, might be more conscientious in the completion the SIS than would a random sample of interviewers. In future reliability studies on the SIS, it would be beneficial if a state or agency selected interviewers at random so that selection bias can be ruled out as a confounding factor.

References

- American Association on Intellectual and Developmental Disabilities (AAIDD). (2007). *SIS Vantage: Perspectives on the Supports Intensity Scale*. Volumes 1 & 2. Retrieved August 31, 2007, from <http://www.siswebsite.org/page.ww?section=Newsletters&name=Past+Issues>
- Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency, 86*, 127–137.
- Clay-Adkins, S. L. (2004). *Reliability and validity of the Supports Intensity Scale*. Unpublished doctoral dissertation, Illinois State University, Normal, IL.
- Eliason, M., LeVelle, J., McLaughlin, C. A., Schalock, R. L., Thompson, J. R., & Tribble, A. (2005). *Supports Intensity Scale expanded item descriptions*. Washington, DC: American Association on Mental Retardation. Retrieved August 31, 2007, from http://www.siswebsite.org/cs/root/main/supports_and_sis/reading_material
- Greenspan, S. (2006). Mental retardation in the real world: Why the AAMR definition is not there yet. In H. N. Switzky & S. Greenspan (Eds.), *What is mental retardation: Ideas for an evolving disability in the 21st century* (rev. and updated). Washington, DC: American Association on Mental Retardation.
- Harries, J., Guscia, R., Kirby, N., Nettelbeck, T., & Taplin, J. (2006). Support needs and adaptive behaviors. *American Journal on Mental Retardation, 110*, 393–404.
- Joint Committee on Standards for Educational and Psychological Testing of the American Education Research Association, the American Psychological Association, and the National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Luckasson, R., Borthwick-Duffy, S., Buntinx, W. H. E., Coulter, D. L., Craig, E. M., Reeve, A., Schalock, R. L., Snell, M. E., Spitalnik, D. M., Spreat, S., & Tassé, M. J. (2002). *Mental retardation: Definition, classification, and systems of supports* (10th ed.). Washington, DC: American Association on Mental Retardation.
- Smith, G., & Fortune, J. (2006). *Assessment instruments and community services rate determination: Review and analysis*. (Available from the Human Services Research Institute, 7420 SW Bridgeport Rd., Ste. 210, Portland, OR, 97224).
- Stancliffe, R. J., & Lakin, K. C. (Eds.). (2005). *Costs and outcomes of community services for people with intellectual disabilities*. Baltimore: Brookes.
- Thompson, J. R., Bryant, B. R., Campbell, E. M., Craig, E. M., Hughes, C. M., Rotholz, D. A., Schalock, R., Silverman, W., Tassé, M., & Wehmeyer, M. (2004a). *Supports Intensity Scale (SIS)*. Washington, DC: American Association on Mental Retardation.
- Thompson, J. R., Bryant, B. R., Campbell, E. M.,

SIS interrater reliability

J. R. Thompson, M. J. Tassé, and C. A. McLaughlin

Craig, E. M., Hughes, C. M., Rotholz, D. A., Schalock, R., Silverman, W., Tassé, M., & Wehmeyer, M. (2004b). *The Supports Intensity Scale (SIS): Users manual*. Washington, DC: American Association on Mental Retardation.

World Health Organization. (2001). *International classification of functioning, disability, and health (ICF)*. Geneva: Author.

Received 5/18/07, accepted 9/11/07.

Editor-in-charge: William E. MacLean, Jr.

The authors thank the following organizations for their willingness to collect data for this study: Arizona Training & Evaluation Center (Peoria, AZ), Centre de Readaptation Lisette-Dupras (Montréal,

QC, Canada), Horizon House (Peru, IL), Lancaster County Mental Health/Mental Retardation Program (Lancaster, PA), Little Friends, Inc. (Naperville, IL), Minda, Inc. (Brighton, South Australia, Australia), Oklahoma Department of Human Services (Oklahoma City, OK), Ray Graham Association (Downers Grove, IL), Services de Readaptation L'Integrale (Montréal, QC, Canada), Shore Training Center (Morton Grove, IL), Temple University, Institute on Disabilities (Philadelphia, PA), Trinity Services (Joliet, IL), West Montréal Readaptation Center (Montréal, QC, Canada). Requests for reprints should be sent to James R. Thompson, Illinois State University, Special Education, Normal, IL 61790. E-mail: jrthomp@ilstu.edu