

# 2011 Validity and Reliability Results Regarding the SIS



---

*Jon Fortune, Ed.D., John, Agosta, Ph.D.,  
and Julie Bershadsky, Ph.D.*

*March 14, 2011*

**Face Validity.** Developed to measure the construct of supports, the SIS has greater face validity than the ICAP or other traditional assessments. The assessment of support needs using the SIS is done directly by persons with first-hand knowledge of the individual. The SIS directly measures the level of supports needed to enable an individual to participate successfully in the life of his or her community. It necessarily looks at more than skills and deficits, considering motivation, health, etiology, problem behavior, environment and other variables influencing the need for supports. By measuring individual support needs directly, it avoids the error inherent in inferring support needs statistically based on adaptive and maladaptive behavior scales. It is transparent. The SIS assessment of needed supports is more explicit and straightforward than other traditional instruments, and hence is a more open platform for the stakeholder deliberation and decision-making that attends individual resource allocation and payment processes. The SIS uses multi-point scales to rate the type (monitoring – full physical assistance), frequency (none to hourly) and intensity (no time to more than 4 hours in a 24 hour period) of supports needed by an individual to participate in 57 distinct aspects of life in their communities. Behavioral, health and other factors affecting support needs are considered.<sup>12</sup> Buros Institute<sup>3</sup> review of the SIS notes: “The developers of the SIS encourage users to aggregate data to use in program planning, resource allocation, funding analyses, and program evaluation. Although this is not the primary purpose of the SIS, it may be used to assist decision makers in those areas.”

**Content Validity.** To assure its content validity, the SIS was constructs were tested by 74 professionals working in the field of developmental disabilities. Using a Q-sort methodology, they narrowed the 130 candidate support indicators to 57, and reduced the 12 domains containing these indicators to seven. This makes the instrument more concise while still asking

---

<sup>1</sup> Thompson, J. R., Bryant, B. R., Campbell, E. M., Craig, E. M., Hughes, C. M., Rotholz, D. A., Shalock, R. L., Silverman, W. P., Tassé, M. J., & Wehmeyer, M. L. (2004). *Supports Intensity Scale. Users' manual*. Washington, DC: American Association on Mental Retardation.

<sup>2</sup> Thompson, J. R., McGrew, K.S., & Bruininks, R. H. (2002). Pieces of the puzzle: Measuring the personal competence and support needs of persons with intellectual disabilities. *Peabody Journal of Education*, 77, 23-39.

<sup>3</sup> Loew, S.A. & Pittenger, D, J. (2005). Test review of the Supports Intensity Scale. In R. A. Spies & B. S. Plake (Eds.), *The sixteenth mental measurements yearbook*. Retrieved from the Buros Institute's *Test Reviews Online* website: <http://www.unl.edu/buros>

the right questions. The validity of the SIS has been examined in a number of countries.<sup>456789</sup> Efforts have been made to see the efficacy of the SIS in predicting extraordinary support needs (N=274).<sup>10</sup> The SIS is often used to inform interdisciplinary team individual service plans<sup>11</sup> and is increasingly used to form resource allocation systems.<sup>12</sup>

**Construct and criterion validity.** The high correlation of SIS subscale scores with one another shows that the SIS measure has good construct validity, meaning that scores on the SIS are highly correlated with scores on measures of other constructs (for example, adaptive behavior and intelligence) that are believed to be correlated with the construct measured by the SIS. To establish its criterion validity, the SIS measures of support needs were correlated with an independently constructed “criterion measure” - a Likert-type scale of support needs. All correlation coefficients exceeded the .35 minimum level required to demonstrate criterion-related validity.<sup>13</sup> Support for the construct validity of the Supports Intensity Scale based on clinician rankings of need (N=50) was explored in Ontario Canada in 2009.<sup>14</sup> Factor analysis in The Netherlands and Belgium with 14,862 individuals has added support to the overall integrity of the SIS and to the confirmatory factor analysis supporting the originally proposed subscale structure like the SIS results like the Section 1 subscales A, B, and E

---

<sup>4</sup> Shalock, R. L., Thompson, J. R., & Tassé, M. J. (2008a). *International implementation of the Supports Intensity Scale*. Washington, DC: American Association on Intellectual and Developmental Disabilities.

<sup>5</sup> Giné, C. (2008). Catalan translation of the Supports Intensity Scale. In R. L. Shalock, J. R. Thompson, & M. J. Tasse (Eds.), *International implementation of the Supports Intensity Scale* (pp. 7-8). Washington, DC: American Association on Intellectual and Developmental Disabilities.

<sup>6</sup> Buntinx, W. H. E., (2008). The Dutch version of the Supports Intensity Scale. In R. L. Schalock, J. R. Thompson, & M. J. Tassé (Eds.), *Psychometric properties of the Supports Intensity Scale* (pp. 6-10). Washington, DC: American Association on Intellectual and Developmental Disabilities.

<sup>7</sup> Llewellyn, G., Riches, V. C., Hindmarsh, G. J., Parmenter, T. R., & Chan, J. (2005). *I-CAN: Instrument to classify support needs for people with disability*. Sydney, Australia: University of Sydney and Centre for Developmental Disabilities Studies.

<sup>8</sup> Morin, D., & Cobigo, V. (2008). The French Version of the Supports Intensity Scale. In R. L. Shalock, J. R. Thompson, & M. J. Tassé (Eds.), *Psychometric properties of the Supports Intensity Scale* (pp. 3-4). Washington, DC: American Association on Intellectual and Developmental Disabilities.

<sup>9</sup> Verdugo, M., Arias, B., Ibanéz, A., & Gómez, L. (2006). *Validation of the Spanish version of the Supports Intensity Scale*. *Journal of Applied Research in Intellectual Disabilities*, 19, 274.

<sup>10</sup> Wehmeyer, M., Chapman, T. E., Little, T.D., Thompson, J. R., Shalock, R., and Tassé, M. J. (2009). Efficacy of the Supports Intensity Scale (SIS) to Predict Extraordinary Support Needs. *American Journal of Intellectual and Developmental Disabilities*, 114(1), 3-14.

<sup>11</sup> Shalock, R. L., Thompson, J. R., & Tassé, M. J., (2008b). Relating Supports Intensity Scale information to Individual Service Plans. Washington, DC: American Association on Intellectual and Developmental Disabilities.

<sup>12</sup> Shalock, R. L., Thompson, J. R., & Tassé, M. J., (2008c). Resource allocation and the Supports Intensity Scale: Four papers on issues and approaches. Washington, DC: American Association on Intellectual and Developmental Disabilities.

<sup>13</sup> Hammill, D.D., Brown, L., & Bryant, B.R. (1992). *A consumer guide to tests in print*. Austin, TX: Pro-Ed.

<sup>14</sup> Weiss, J. A., Lunskey, Y., Tassé, M. J., & Durbin, J. (2009). Support for the construct validity of the Supports Intensity Scale based on clinician rankings of need. *Research in Developmental Disabilities*, 30, 933-941.

and the Section 3 Total Medical and Total Behavioral needs.<sup>15</sup> The Spanish study of the SIS suggests similar value of the SIS.<sup>16</sup>

**Internal Consistency.** The SIS is internally consistent.<sup>17,18</sup> It has good inter-item reliability (all items or subscales in the measure are measuring the same construct). The internal consistency reliability coefficients for all the SIS subscales, computed using Cronbach's Alpha method<sup>19</sup>, exceeded .90, which is the level widely accepted as demonstrating an acceptable level of internal consistency in assessment scales.

**Inter-rater Reliability.** The SIS has a high degree of inter-rater reliability<sup>20</sup>: the SIS Index (total score) correlation coefficient was .87 (same interviewer, different respondent), .90 (different interviewer, same respondents), and .85 different interviewer and different respondents (N=40). These more recent results on inter-rate reliability are an improvement over the inter-rate reliability results reported in the SIS manual of .54 for Pearson r and .59 for the corrected r.

The most comprehensive study of the inter-rater reliability was reported in the article by Thompson, Tasse, and McLaughlin in 2008<sup>21</sup>. This is the only study where the authors attempted to parcel out variance from having different respondents (same interviewer interviewing different respondents), different interviewers (different interviewers interviewing the same respondents) and both (different interviewers independently interviewing a different pair of respondents) and found that under all three conditions the inter-rater reliability was excellent based on the Cichetti and Sparrow guidelines for AB scales: "if policymakers have decided that measuring the intensity of a person's support needs provides information useful in the creation of reimbursement formulas or determining individualized resource allocations, they can be confident that the SIS is a reliable measurement tool."

This answers the concern noted earlier by the Buros Institute<sup>22</sup>:

"The retests were given by the same interviewer who had given the first test. The interrater reliability coefficients ranged from .55 to .90, which is disturbing. Given that a case manager in an entry level position would probably be the interviewer using this instrument and then when

---

<sup>15</sup> Kuppens, S., Bossaert, G., Buntinx, W., Molleman, C., Abbeele, A. V. & Maes, B. (2010). Factorial validity of the Supports Intensity Scale (SIS). Factorial Validity of the Supports Intensity Scale (SIS). *American Journal on Intellectual and Developmental Disabilities*, 115(4), 327-339.

<sup>16</sup> Verdugo, M.V., Arias, B, Ibanez, A, & Schalock, R. L. (2010). Adaptation and psychometric properties of the Spanish version of the Supports Intensity Scale (SIS). *American Journal on Intellectual and Developmental Disabilities*, 115(6), 496-503.

<sup>17</sup> Tassé, M. J. Thompson, J. R. & McLaughlin, C. (2006). *Inter-interviewer and inter-respondent concordance on the Supports Intensity Scale*. Poster presentation at the International Summit for the Alliance on Social Inclusion. May 3-5. Montreal, Canada.

<sup>18</sup> Thompson, J. R., Tassé, M. J., & McLaughlin, C. A. (2008). *Interrater reliability of the Supports Intensity Scale (SIS)*. *American Journal on Mental Retardation*, 113, 231-237.

<sup>19</sup> Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.

<sup>20</sup> Thompson, J. ( Feb 21, 2006). SIS reliability: preliminary findings and procedures. Email from J. Thompson to J. Ashbaugh.

<sup>21</sup> Thompson, J. R., Tassé, M. J., & McLaughlin, C. A. (2008). *Interrater reliability of the Supports Intensity Scale (SIS)*. *American Journal on Mental Retardation*, 113, 231-237.

<sup>22</sup> Loew, S.A. & Pittenger, D. J. (2005). Test review of the Supports Intensity Scale. In R. A. Spies & B. S. Plake (Eds.), *The sixteenth mental measurements yearbook*. Retrieved from the Buros Institute's *Test Reviews Online* website: <http://www.unl.edu/buros>

it is time for follow-up or review of Individualized Service Plans, there would probably be a new case manager in that same entry level job, interrater reliability would be very important. Although the psychometric properties of the instrument are adequate for the most part, the most troubling aspect is low interrater reliability. The reality for many agencies is a large turnover in staff, which means that a consumer might be served by the same agency for years but be interviewed by a number of different case managers during those years. The discrepancies that might result from low interrater reliability should be of concern for anyone using this instrument.”

All correlation coefficients exceeded the .35 minimum level required to demonstrate criterion-related validity.<sup>23</sup> Perhaps more importantly, the inter-rater reliability of the SIS is as good or better than most AB scales on the market; so, it is as good as any other alternative one might want to use. The other thing perhaps worth noting is that the interviewers in our study had all gone through training, so it is not too big of a leap to suggest that if one wants a high degree of reliability between different interviewers, then they should make sure interviewers have received comparable training on how to administer the scale.<sup>24</sup>

**Test-Retest Reliability.** The SIS manual suggests a test-retest Pearson r results for the SIS total section 1 results of .79 with .82 corrected r.<sup>25</sup> The manual using the data from Clay study in Illinois of test-retest and inter-rater reliability<sup>26</sup> also provides a detailed table of breakout of the six subscales with Home Living reported as .87, Community Living reported as .74, and Health and Safety reported as .86 (all corrected Pearson r results).

The Buros Institute review<sup>27</sup> of the SIS posits: “Test-retest reliability coefficients (approximately 3-weeks period) ranged from .74 to .94, which is within an acceptable range. The retests were given by the same interviewer who had given the first test.” “The test-retest (with a 3-week administration interval) and interrater reliability indices indicate a high level of test-retest reliability, but only moderate levels of interrater reliability. The latter observation is not surprising given the nature of the instrument and the methods for collecting the information. The interrater reliabilities vary considerably among the subscales. For example, the Home and Living, and Health and Safety subscales evidence relatively robust interrater reliabilities. By contrast, the Lifelong Learning and Employment subscales evidence notably lower levels of interrater reliability.”

---

<sup>23</sup> Hammill, D.D., Brown, L., & Bryant, B.R. (1992). *A consumer guide to tests in print*. Austin, TX: Pro-Ed.

<sup>24</sup> Thompson, J. (Mar 13, 2011). SIS reliability: updated findings. Email from J. Thompson to J. Fortune.

<sup>25</sup> Thompson, J. R., Bryant, B. R., Campbell, E. M., Craig, E. M., Hughes, C. M., Rotholz, D. A., Shallock, R. L., Silverman, W. P., Tassé, M. J., & Wehmeyer, M. L. (2004). *Supports Intensity Scale. Users' manual*. Washington, DC: American Association on Mental Retardation.

<sup>26</sup> Clay, S. (2003). The reliability and validity of the AAMR Supports Intensity Scale. Unpublished study. Normal: Illinois, Department of Special Education.

<sup>27</sup> Loew, S.A. & Pittenger, D, J. (2005). Test review of the Supports Intensity Scale. In R. A. Spies & B. S. Plake (Eds.), *The sixteenth mental measurements yearbook*. Retrieved from the Buros Institute's *Test Reviews Online* website: <http://www.unl.edu/buros>

The Dutch two pilot studies<sup>28</sup> and Spanish pilot study<sup>29</sup> mirror the validity and reliability of the SIS corresponding with the originally published psychometric properties. The most recently published work which is from Spain goes on to describe the test-retest reliability of the SIS in a particularly illustrative manner:

“The same interviewer applied the SIS to a sample of 143 participants at two different times with an interval of 3 weeks. We calculated the Pearson’s product moment correlation coefficients between the test and retest scores and obtained values between .84 and .93 (between .901 and .981 in the corrected correlations). These coefficients are excellent according to Cicchetti and Sparrow<sup>30</sup>. Specifically, the test-retest coefficients for each of the 49 items ranged from .63 (Item 4 of Health and Safety Activities, ambulating and moving about) to .90 (Item 3 of Home Living Activities, preparing food). Using the guidelines provided by Cicchetti and Sparrow for evaluating the reliability coefficients, we found that 46 coefficients (.93.9%) fell within the excellent range, and only 3 (6.12%) in the good range, which guaranteed the test-retest reliability of the SIS.”

**Summary of the SIS by the Buros Institute<sup>31</sup>:** “The SIS is a valuable tool to assist agencies in the development of Individualized Service Plans. It does not take the place of diagnostic tests but is used after a diagnosis is made to determine the supports needed and the intensity of those supports for a person to be a functioning member of the community. The SIS is well-researched and easy to use and score with a comprehensive manual that is easy to read and use. Because it is such a useful and well-constructed instrument, it is essential for any agency that chooses to use the SIS to conduct regular training in the use and scoring of the instrument so that interrater reliability within the agency is in acceptable ranges. The developers encourage users to share the results of using the SIS with different samples so that information can be included in subsequent manuals. It is clear that they recognize that although they have made an excellent start, there is still room for improvement. The SIS is a carefully designed semistructured interview instrument that allows persons with limited training and supervision to assess quickly the level of functioning and needs of a person with developmental disabilities. The construction of the instrument reveals careful attention to detail and the assurance that it produces useful information. Moreover, the instrument is sufficiently flexible to ensure its use in a number of settings.”

---

<sup>28</sup> Kuppens, S., Bossaert, G., Buntinx, W., Molleman, C., Abbeele, A. V. & Maes, B. (2010). Factorial validity of the Supports Intensity Scale (SIS). Factorial Validity of the Supports Intensity Scale (SIS). *American Journal on Intellectual and Developmental Disabilities*, 115(4), 327-339.

<sup>29</sup> Verdugo, M.V., Arias, B, Ibanez, A, & Schalock, R. L. (2010). Adaptation and psychometric properties of the Spanish version of the Supports Intensity Scale (SIS). *American Journal on Intellectual and Developmental Disabilities*, 115(6), 496-503.

<sup>30</sup> Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *American Journal of Mental Deficiency*, 86, 127-137.

<sup>31</sup> Loew, S.A. & Pittenger, D, J. (2005). Test review of the Supports Intensity Scale. In R. A. Spies & B. S. Plake (Eds.), *The sixteenth mental measurements yearbook*. Retrieved from the Buros Institute's *Test Reviews Online* website: <http://www.unl.edu/buros>