

MAY 2018

Evidence for the Reliability and Validity of the Supports Intensity Scales

Authors: James R. Thompson, Robert L. Schalock, and Marc J. Tassé



American Association
on Intellectual and
Developmental Disabilities

www.aaidd.org

Authors' Notes

Research findings published in peer-reviewed professional literature in regard to the Supports Intensity Scale—Adult Version (SIS—A) and Supports Intensity Scale—Children's Version (SIS—C) are reviewed in this white paper. Background information regarding the rationales for and processes used in developing the SIS—A and SIS—C is provided. Core concepts associated with psychological measurement are briefly explained, and research findings on the SIS—A and SIS—C are summarized in relation to these core concepts.

I. Creating a New Scale to Assess the Support Needs of People With Intellectual Disability

The first American Association on Intellectual and Developmental Disabilities manual on terminology, definition, and classification of intellectual disability was published in 1910. The committee that wrote it presented it to the AAIDD board in the form of a report (Rogers, 1910), and they could not have foreseen that their work would be a first edition of a manual that would have a significant influence on how people with intellectual disability and related developmental disabilities were understood over the next 100 years and beyond. Approximately every decade or so, a new committee is formed by AAIDD to examine the prior edition of the association's manual and update it based on knowledge obtained from new research findings, as well as changing needs and practices in the field of human services for people with disabilities.

The 9th edition of *Intellectual Disability: Definition, Classification, and Systems of Supports* (Luckasson et al., 1992) was particularly transformative. The authors called for shifting the conceptualization of intellectual disability from a deficit model to a social-ecological model. According to the 9th edition, intellectual disability was better understood as a “state of functioning” evidenced by a person's competencies not being well aligned with the environmental demands, in contrast to the traditional view that it was a deficit trait within an individual.

Wehmeyer et al. (2008) pointed out that the 1992 manual and subsequent editions (i.e., Luckasson et al., 2002; Schalock et al., 2010) provided two definitions of intellectual disability: an operational definition that “operationalizes the intellectual disability construct and provides the basis for diagnosis and classification” (p. 311) and a constitutive definition that “explains the underlying construct and provides the basis for theory-model development and planning individualized supports” (p. 311). The operational definition makes the AAIDD manual workable for practicing diagnosticians by requiring documentation of relative deficits in intellectual functioning and adaptive behavior skills. The constitutive definition, in contrast, is centered on a contextual understanding of people and the environments in which they live and interact. According to the constitutive definition of intellectual disability, the most salient difference between people with intellectual disability and the general population is that people with disabilities need different types and intensities of support in order to fully participate in and contribute to the settings and activities of daily life.

Proponents from multiple disciplines have perceived the concept of understanding people with intellectual disability by their needs for support to be more useful for purposes of planning than conventional deficit-based conceptualizations (see Fredericks & Williams, 1998; Ward & Stewart, 2008). However, soon after the publication of the 9th edition of AAIDD's manual, it became clear that a means to accurately and objectively

measure support needs was needed. Without a reliable and valid process to measure it, “support needs” was an ethereal construct that could be described and understood in so many different ways that it was at risk of becoming meaningless for purposes of both research and application. Relying on informal, nonuniform characterizations of various kinds of support that different people needed made it difficult to pinpoint how the support needs of people with intellectual disability differed from those of the general population. Also, because people with intellectual disability are a heterogeneous population, the absence of a uniform (and defensible) means to assess support needs made distinctions among people with intellectual disability an imprecise undertaking.

Recognizing that the means to accurately measure a construct of interest often results in important innovations and advances in any field of research or practice, the AAIDD charged a team of researchers with creating a psychometrically reliable and valid assessment tool to measure the pattern and intensity of people’s support needs. The outcome of their work was the Supports Intensity Scale, which is now referred to as the Supports Intensity Scale—Adult Version (SIS—A; Thompson et al., 2015) in order to distinguish it from the Supports Intensity Scale—Children’s Version (SIS—C; Thompson et al., 2016). As mentioned, the 9th edition of AAIDD’s manual served as a catalyst for developing the SIS—A. But, the authors of the original *Supports Intensity Scale (SIS): User’s Manual* (Thompson et al., 2004) also identified five trends in the field of intellectual and developmental disabilities that led to interest in developing ways to better understand people by their needs for support and creating a standardized assessment scale to measure people’s support needs. These five trends, which began in the early 1960s and continue today, bear revisiting:

- **Changes in expectations for people with disabilities.** Sixty years ago, many life experiences (e.g., attending school in general education classrooms alongside same-age peers, living in a home of one’s choice, holding a paid job, having a long-term romantic relationship) were perceived to be unrealistic and, therefore, unattainable for children and adults with intellectual and developmental disabilities. Today, it is not only realistic to expect people with disabilities to live their lives as full-fledged members of their communities, it is perceived to be a failure of the system when individuals are relegated to society’s margins and not provided opportunities to be engaged in culturally valued life experiences. Changing expectations called for developing new ways to assess people’s support needs that were relevant to identifying supports that lead to full participation in community life.
- **Functional descriptions of disabilities.** Sixty years ago, biological descriptions (e.g., Down syndrome, phenylketonuria) or severity-based IQ descriptions (e.g., borderline, mild, moderate, severe, profound) dominated the ways in which people with intellectual disability and related developmental disabilities were described. Such descriptions provided limited guidance in terms of supporting children and adults with intellectual disability to reach meaningful life goals. In contrast, functional descriptions that focused on how a person actually functioned/operated/performed in daily life activities in contemporary society provided a basis on which to identify skills that would be useful for a person to learn, identify tools (i.e., assistive technologies) that an individual might use to enhance their participation in settings and activities, and/or modify the design or the demands of settings and

activities so that a person could be accommodated. Teaching skills, incorporating technological solutions, and modifying environments are all included under the umbrella of personalized supports because their purpose is to decrease the mismatch between personal competencies and environmental demands. New ways to assess people's support needs were needed to develop new types of functional descriptions that could better inform the identification and coordination of support strategies.

- **Chronologically age-appropriate activities.** At one time, it was assumed that people with intellectual disability had minds similar to those of children and, therefore, life activities and experiences were aligned with people's "mental ages." Due to their perceived childlike state and vulnerabilities, caregivers felt justified in controlling every aspect of people's lives throughout the lifespan. Paternalistic attitudes deprived people of opportunities to take the types of risks associated with becoming an adult member of society. Furthermore, the "perpetual child" status in society reduced the respect that was shown to them by others. The movement toward ensuring access to chronologically age-appropriate settings and activities shifted the focus from how to protect people from society's threats to how to integrate people meaningfully into society's fabric. New ways to assess people's support needs were needed that yielded insights into ways in which people with intellectual disability could be supported to assume meaningful, age-appropriate roles and identities in childhood (e.g., student, friend) and adulthood (e.g., employee, neighbor).
- **Consumer-driven services and supports.** Using public resources to support programs and services targeted to people with intellectual and developmental disabilities in local communities (as opposed to large institutions) was a seismic change in public policy during the 1950s and 1960s. Although establishing local community service systems provided new opportunities to people with intellectual and developmental disabilities and their families (i.e., the consumers of the services), people had to conform to the programs that were offered in order to receive assistance. For example, if a person wanted assistance in finding and keeping a job, they needed to access services offered by a local provider organization's vocational program. Consumer-driven services and supports flipped this equation by requiring service providers to tailor their efforts to the needs and preferences of individual people. That is, instead of offering programs that people with intellectual disability could take or leave, the new approach called for people to indicate how they wanted to be supported, and it was the community service systems' responsibility to tailor supports to people's needs and preferences. Although consumer-driven services and supports are intertwined with several other trends and concepts (e.g., person-centered planning, self-determination, shared decision making, individualized budgeting), the importance of transforming a community-based service system from one that was designed to offer "programs" across broad life domains (e.g., vocational, residential, recreational) to one designed to be responsive to individual needs and preferences cannot be overstated. New ways to assess people's support needs were needed to transform a system founded on "group-based programs" to a system based on "personalized supports."
- **Support networks that provide individualized supports.** This final trend reflects the identity and role of those who provide direct supports to people with disabilities. In the past, it was assumed that people with disabilities needed caregivers to assist with

activities of daily living, such as dressing, eating, and maintaining safety. Although some people with disabilities require “hands-on” personal assistance for parts of their day, an expanded vision for those providing support to people with disabilities has emerged that goes well beyond that of a “caregiver” who meets people’s basic physical needs and ensures their safety. People with disabilities need access to *facilitators of community integration and participation*, as opposed to caregivers who perceived themselves as having a narrow purpose. Facilitators of community integration and participation are charged with assisting people with disabilities in establishing support networks comprised of many individuals who provide many different types of support. Support networks might include natural supports (i.e., supports that are preexisting in the environment, such as family members, friends, coworkers, neighbors, classmates, bus drivers, and police officers), paid staff, and others. Good support networks take considerable time and effort to establish. Maintaining support networks requires ongoing efforts, as people’s support needs change over time and environmental demands change as well. New ways of assessing people’s support needs were needed to inform the development and maintenance of responsive and strong support networks.

II. Evolving Understandings of Supports and Support Needs, and the Need for Psychometric Evidence

Conceptual Foundations and Essential Terminology

Since the publication of the original Supports Intensity Scale in 2004 (Thompson et al.), considerable conceptual and empirical work has been completed regarding the nature of support needs and supports. Support needs were defined as “a psychological construct referring to the pattern and intensity of supports necessary for a person to participate in activities linked with normative human functioning” (Thompson et al., 2009, p. 135). Like other psychological constructs (e.g., anxiety, intelligence, happiness, morality), people can experience extreme points. For example, in regard to happiness, people can be euphoric and depressed, but there are many points in between. Similarly, people can have very high intensity or very low intensity support needs, with a continuum of intensities between the two extremes. A person’s intensity of support needs is a reflection of the extent of congruence between individual capacity and the environments in which an individual is expected to function. Multiple factors can influence the extent of the mismatch, including personal competency related to various domains of functioning (e.g., physical competency, conceptual competency, practical competency, social competency, emotional competency) and environmental demands related to the complexity of settings and activities (Thompson, Shogren, & Wehmeyer, 2017).

Supports are defined as “resources and strategies that aim to promote the development, education, interests, and personal well-being of a person and that enhance individual functioning” (Luckasson et al., 2002, p. 151). The purpose of supports is to bridge the gap between the person and the environment and, therefore, supports are evidenced by the extent to which a person’s participation and engagement in settings and activities is enhanced after supports are introduced. Just because something is *intended* to be a support does not mean it actually *is* a support based on Luckasson’s et al. definition. That is, supports are evidenced by results.

For example, if a person is provided vocational services in a vocational center, but these services never lead to a paid job in the community, the vocational services this person receives do not function as a support in regard to competitive employment. This is not to suggest the person did not benefit from the vocational services that were received (e.g., perhaps the person enjoyed their time in the vocational center and expressed satisfaction with the center’s services). It simply reflects the reality that there was no evidence to suggest the gap between the person and the demands of the environment (i.e., the settings and activities of competitive employment) was successfully bridged as the result of services provided at the vocational center. Likewise, a paraprofessional who serves as a 1:1 aide for a child throughout the school day and works with a child on activities and assignments that are quite different from those in which other children in the general

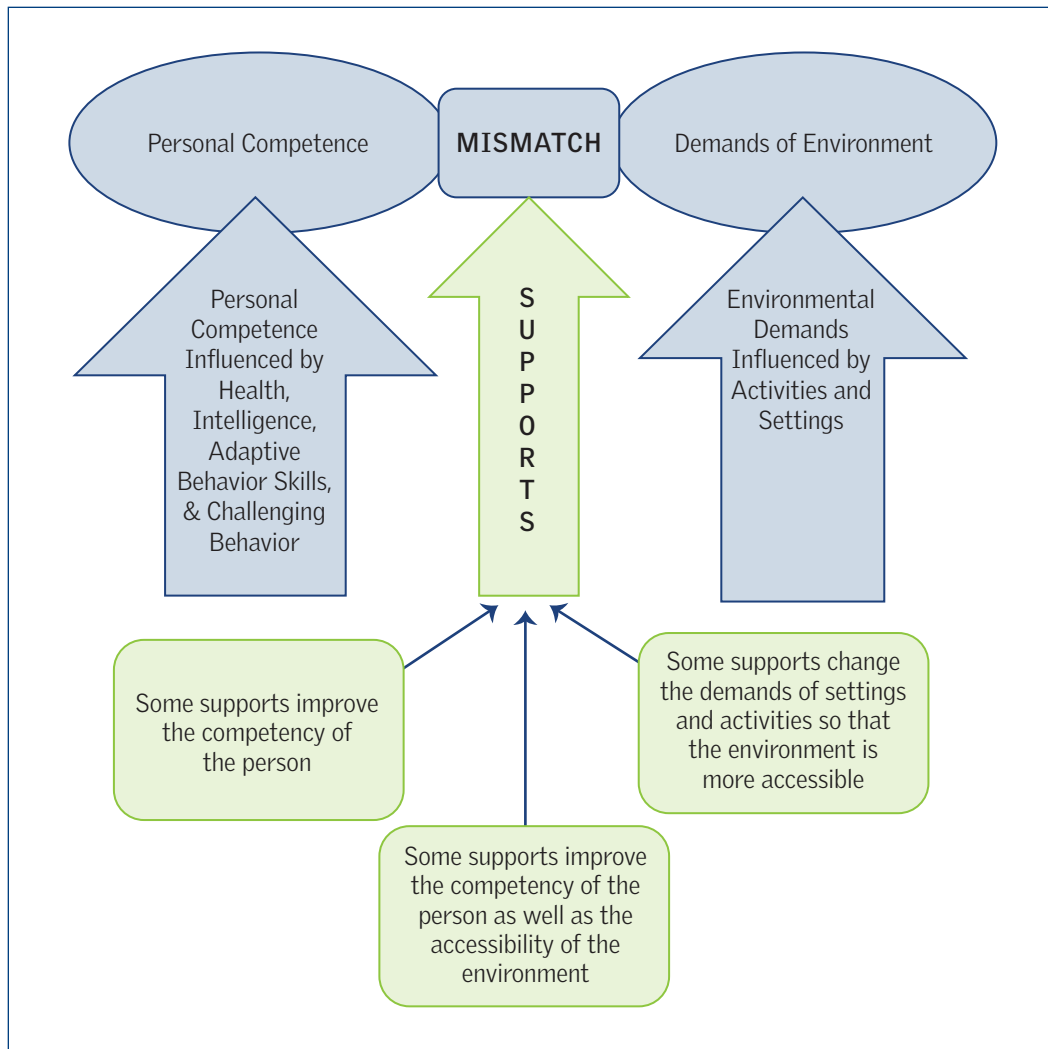
education classroom are engaged is only addressing the gap between the student and classroom settings/activities superficially. The paraprofessional's support is maintaining the physical presence of the child in the classroom, but it is not bridging the student-environment mismatches relevant to participation in classroom learning activities.

Unique Information Provided by the SIS—A and SIS—C

Figure 1 illustrates the relationship between support needs and supports, and ways in which supports can function to improve the person's competency and/or change environmental demands to address the person-environment mismatch. There are numerous assessment instruments in the field of intellectual and developmental disabilities (IDD) to measure aspects of personal competence (e.g., IQ tests measure intelligence and adaptive behavioral scales measure adaptive skills). Also, there are well-established techniques to assess performance demands associated with specific environments, such as the ecological inventory (Sobsey, 1987) and the job analysis (Clark & Kolstoe, 1995). Assessment instruments and processes to measure personal competence and analyze performance requirements in various settings continue to be useful, and the SIS—A and SIS—C were not designed to displace them. However, traditional assessments to measure personal competence and environmental demands called for users who were interested in understanding people by their support needs, to make inferences from different sources of data that were not designed to be aligned with one another. The SIS—A and SIS—C were developed to provide a means to directly measure the extent of the person-environment mismatch shown at the top of Figure 1.

As mentioned above, there has been considerable interest in supports needs assessment and planning since the original Supports Intensity Scale was published nearly 15 years ago. The SIS—A and SIS—C are currently being used throughout North America and internationally (see American Association on Intellectual and Developmental Disabilities, 2017a; 2017b, for listings of jurisdictions where these scales are being used). Also, several companion products have been developed in response to user requests. A workbook providing detailed, step-by-step guidelines for planning teams that wish to incorporate SIS—A results within a person-centered planning process has been published (Thompson, Doepke, Holmes, Pratt, Myles, Shogren, & Wehmeyer, 2017). Moreover, the *SIS—A Annual Review Protocol* (Thompson, Shogren, Schalock, Tassé, & Wehmeyer, 2017) was recently published to provide planning teams with a tool to inform discussion regarding a person's need for reassessment with the SIS—A when a prior assessment has been administered. Additionally, a user's guide, the *SIS—A Annual Review Protocol: A Facilitator's Guide* (Thompson, Shogren, Wehmeyer, Schalock, & Tassé, in press), is currently in production.

FIGURE 1
Supports address the mismatch between people and environments.



Establishing the Technical Adequacy of the SIS Scales

The adoption of the SIS—A and SIS—C by numerous jurisdictions and the demand for a suite of products associated with these assessment scales reflects the motivation of policy makers, professionals, and consumer advocates to embrace a social-ecological understanding of disability and create a progressive system of service delivery where consumer choices and support needs are the driving influences. It is also a reflection of the confidence people have in the quality of the two scales. Research findings regarding the technical properties of the SIS—A and SIS—C provide the foundation for justifying the use of the assessment results for purposes of decision making, whether decisions are made at micro (individual), meso (organizational), or macro (jurisdictional) levels (Thompson, Schalock, Agosta, Teninty, & Fortune, 2014). In the next section, concepts that are essential to evaluating the psychometric quality of assessment instruments

are briefly explained. These same concepts provide the basis for the summarization of findings from the professional literature on the psychometric properties of the SIS—A and SIS—C that is presented in the final section of this white paper.

Reliability

Assessment tools must be reliable and valid, and there are numerous ways to evaluate their attributes in regard to these properties. *Reliability* refers to the degree that an assessment measures a construct consistently. *Internal consistency reliability* refers to the extent to which the items correlate with one another. Strong correlations indicate that an assessment scale's items have a robust relationship with one another and are, therefore, measuring different aspects of the same construct. *Split-half reliability* refers to the linear relationship between half of the items on a scale with the other half. Items on an assessment scale are randomly divided into two sets. The entire instrument is administered to a sample of people and scores on both halves are compared. A high correlation between the two sets of scores suggests that items on the scale are measuring the same construct. *Test-retest reliability* evaluates the consistency of a scale score over short periods of time. If the construct being measured is conceptualized as being stable (i.e., not subject to a significant change over a short period of time), strong correlations should be apparent between scores from two separate and independent administrations completed under the same conditions (e.g., same assessor) but at different time points. *Interrater reliability* refers to the consistency of scale scores across assessors (i.e., interviewers or respondents in the case of the SIS scales). When two separate and independent administrations of an assessment involve different assessors, highly correlated scores from the two administrations would suggest that assessment results were trustworthy regardless of who was administering the scale.

Validity

Validity refers to the extent to which an assessment measures what it purports to measure. Although reliability is a necessary condition for establishing the validity of a scale, it is also insufficient. Simply because a scale provides a reliable measure of support needs does not mean it provides a valid measure of support needs. Just as reliability is not established by a single indicator, there are multiple methods used to accumulate evidence of an assessment scale's validity.

Content validity refers to the extent to which items on an assessment accurately represent the universe of items that could be associated with the construct of interest. Assessment developers should attempt to establish content validity at the time assessment is developed, when subscales are conceptualized and items are written. Evidence for content validity includes a record of the process for developing items, including that multiple perspectives were solicited to confirm that the items were reasonable and sources documenting prior knowledge (e.g., peer-reviewed journal articles) were comprehensively reviewed. In addition to documenting the process by which a scale was created, content validity can be evaluated through item analysis. The relationship between each item and the other items on an assessment scale should be examined, as each item should contribute something unique from other items.

Criterion-related validity refers to the extent to which scores on an assessment are correlated with a criterion measure, such as scores on another assessment that measures the same construct. Very low correlations would suggest that the new measure is not measuring the same construct as the established criterion measure. On the other hand, exceptionally high correlations would suggest that the two measures were highly redundant and it would be unlikely that the new measure provide new information or insights. However, if the newer scale had certain advantages that the established criterion measure lacked (e.g., shorter time to complete the assessment, computer administration), then content redundancy may not be a reason for concern.

Construct validity is the most important type of validity. It reflects the extent to which there is converging evidence showing that an assessment truly measures a theoretical characteristic or concept. To evaluate construct validity, Salvia and Yssledyke (2010) recommended examining evidence related to: (a) *internal structure*—items on a scale should share variance in a manner that suggests a common construct is being measured and, when there are subscales, there should be evidence for a multidimensional latent-factor structure; (b) *convergent and discriminant power*—relationships, grounded in theory, should be hypothesized and tested (e.g., because relative intensity of support needs would be hypothesized to have a reciprocal relationship with relative strengths in personal competence, support needs intensity measures should have a negative correlation with measures of intelligence); and (c) *the consequences of testing (predictive)*—an assessment should differentiate between groups of people known to differ on the construct of interest, such as people with less intense support needs (e.g., people with no health problems) and people who have relatively more intense support needs (e.g., people with chronic and serious health problems).

Finally, the concept of *external validity*, although traditionally applied to experimental studies, is also applicable to evaluating the construct validity of assessment scales. External validity refers to the extent to which results from a study can be generalized to other settings and people. Applying this concept to investigating the validity of assessment scales, it is important to evaluate the extent to which an assessment scale's psychometric properties remain robust when applied to different disability groups and/or translated into different languages and used with different cultures.

III. Evidence of the Psychometric Properties of the SIS—A and SIS—C in Peer-Reviewed, Professional Literature

Search Procedures

The following electronic databases were searched using the search phrase “Supports Intensity Scale”: Social Sciences Citation Index, MEDLINE/PubMed, PsycInfo and ERIC databases. The quotation marks were included in the search so that Supports Intensity Scale would be searched as a consecutive string of words, and not as separate words that could appear anywhere within a resource. To be selected from the search, an article had to have been written in English (Note: the Soltani, Kamali, Chaobk, and Ashayeri (2013) article was published in Persian, but the authors provided specific information in their English abstract regarding their data collection procedures and findings. Because their level of detail was comparable to what was presented in the body of the manuscripts of other articles investigating similar research questions, the findings from their article was included in this review.), published in a peer-reviewed professional journal between January 1, 2002 (the first article on the Supports Intensity Scale was published in 2002) and December 31, 2017, and provide results from data analysis pertaining to the psychometric properties of the SIS—A or the SIS—C.

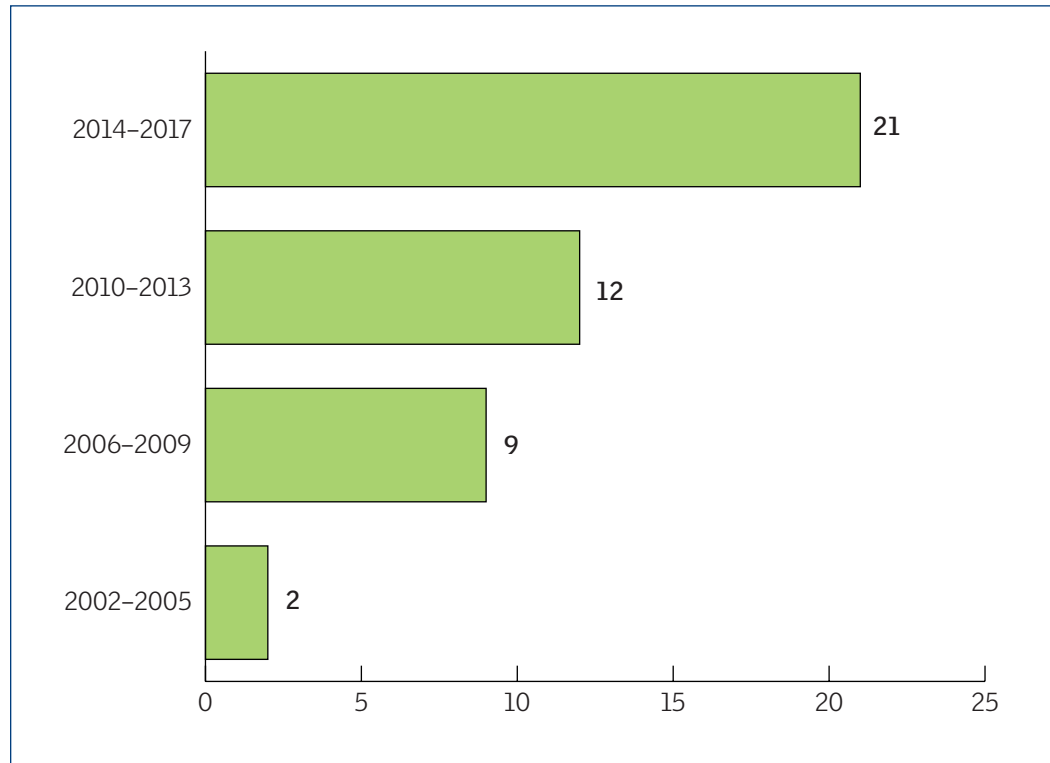
It is important to acknowledge that there has been a considerable amount of grey literature (i.e., literature that is published outside of professional journals and is not subject to peer review) that has focused on the SIS—A and SIS—C over the past 16 years. Many of these publications provide very useful information on the SIS scales. For example, the Human Services Research Institute has produced multiple high-quality publications analyzing SIS data in relationship to a variety of outcome and funding measures (see <https://www.hsri.org/publications>). Additionally, both the SIS—A and SIS—C user’s manuals are part of the grey literature, and both include comprehensive findings from data analyses that are relevant to the psychometric properties of the scales.

Despite the valuable information in the grey literature, we chose to limit our review to publications in peer-reviewed professional journals for two primary reason. First, the volume and diversity of grey literature, which includes online and print jurisdictional reports and guidelines, makes it difficult to ensure that sources are not overlooked. Second, the quality control that is inherent to peer-reviewed literature was important for the purpose of this review. To gain an accurate assessment of the collective research evidence pertaining to the reliability and validity of the SIS scales, it was essential that the research on which the review is based had been evaluated for quality (through peer review) before publication. Peer review provides a foundational level of assurance that the research methods and corresponding findings are scientifically defensible. Therefore, all the studies included in this review were examined by other researchers and judged to possess sufficient scientific rigor for publication in a professional journal.

The search results initially generated a listing of 129 potential sources from professional journals. Sources were eliminated that (a) were duplicates; (b) reported research findings, but did not report any data collection or data analyses from the SIS—A and/or SIS—C that was relevant to evaluating reliability or validity (e.g., see Walker, DeSpain, Thompson, & Hughes, 2014); (c) were conference proceedings that only included a short abstract of the information presented at a conference session (e.g., see Alonso, Gomez, & Navas, 2012); or (d) only provided a description and/or review of either the scale or its corresponding manual (e.g., see Davison, 2005). After the inclusion and exclusion criteria were applied, 42 sources remained from the online search of the databases. Two additional peer-reviewed articles published in scholarly journals were identified through ancestral searches of the reference sections of the identified articles. In total, findings from 44 articles were considered in this review.

Figure 2 shows the number of peer-reviewed journal articles reporting psychometric findings on either the SIS—A or SIS—C that were published during 4-year time frames. It reveals a steady dissemination of research findings relevant to the psychometric properties of the SIS—A and SIS—C since 2006, with an upward trend in volume over time. Citations for the 44 articles are listed in Table 1, with full references provided in a separate reference section in Appendix A. Table 1 identifies the types of psychometric findings that were reported in each of the studies.

FIGURE 2
Number of publications over time.



Statistical Frameworks to Evaluate Psychometric Properties

Knowledge claims regarding the psychometric properties of the SIS—A and SIS—C can emerge from data analysis procedures that are grounded in classical test theory (CTT) or latent variable methods (e.g., methods associated with confirmatory factor analysis, methods aligned with item response theory). Although a variety of statistical approaches can be used to generate useful information, latent variable methods are considered to be superior to CTT by contemporary experts in psychological measurement. Latent variable methods better account for measurement error because they reflect the relationship of items to latent variables as well as the underlying latent structure of the construct that is measured (see Seo, Little, Shogren, & Lang, 2016).

Useful information can be generated, however, from various statistical frameworks. It is beyond the scope of this white paper to discuss assumptions on which different approaches are based or to offer an analysis regarding the relative merits of one approach over another. What is important for the reader to understand is that all statistical models share one thing in common—findings that are generated from statistical analysis are all based on the laws of probability. A robust indicator of psychometric quality (i.e., a statistic generated from data analysis procedures that indicates an assessment scale is reliable and/or valid) should be understood to reflect the reality that there was only a very small possibility that a finding was due to a chance occurrence.

All statistical frameworks provide a means to produce knowledge claims that are said to have empirical evidence, meaning that knowledge claims emerge from investigating patterns in a set of observations (e.g., from a data file that includes responses to items on the SIS scales) and not from logic or theory (although logic or theory should provide the impetus for collecting data). It is important to remember, however, that no statistical approach can compensate for faulty data. The adage “garbage in, garbage out” is applicable to both CTT and latent variable methods. Although methodologists can debate the merits of different statistical techniques, differences in findings from investigations may have more to do with the quality of data collection than the approach to statistical analyses. Analysis using any statistical framework has the potential to provide insights into an assessment scale’s quality as long as data are collected using the correct procedures and accurately entered into a data file.

Despite the advantages of using latent variable methods to investigate an assessment tool’s psychometric properties, CTT has been the traditional framework for generating psychometric indicators through data analysis. The terms used in Table 1 emerged from this framework. These terms are familiar to a wide audience of SIS—A and SIS—C users. Although analyses using latent modeling procedures address the same psychometric concepts related to reliability and validity as CTT, the statistics emerging from the frameworks can be different. For instance, coefficient alpha provides the measure for internal consistency reliability when using CTT methods, but coefficient omega is generated when latent variable methods are employed.

For the purpose of this review, we did not classify studies based on what statistical framework was used, nor did we categorize investigations according to the particular statistics that were reported. Rather, we classified the studies based on the concepts associated with reliability and validity that were discussed in the previous section. Thus,

although Guillen, Adam, Verdugo, and Giné (2015) reported coefficient omega values, and Arnkelsson and Sigurdsson (2014) reported coefficient alpha values, both research teams investigated internal consistency reliability, and that is what is recorded in Table 1.

Summary of Findings From the Peer-Reviewed Literature

Reliability

Findings in regard to the reliability of the SIS—A and SIS—C reflect the extent to which these assessment scales measure support needs (i.e., the construct of interest) consistently. Indicators of strong reliability provide evidence for replication dependability—if one cannot rely on an assessment to produce consistent results, then one cannot have confidence in the results. The greater the reliability of an assessment, the less assessment results are influenced by random error (i.e., measurement error).

Internal consistency reliability. As mentioned earlier, a strong correlation among items on a scale is an indicator of good internal consistency reliability. In contrast, weak correlation among items is indicative of poor internal consistency reliability, and such findings are concerning because they suggest that items on the scale are measuring different constructs. The internal consistency reliability of the SIS—A was investigated in 12 studies, and the coefficients that were reported in all 12 studies exceeded the .90 level that is traditionally cited as the desired standard for demonstrating adequate reliability (e.g., see Aiken & Groth-Marnat, 2005). Equally strong results were reported in the five studies where SIS—C data were collected.

Split half reliability. Similar to item analysis, the focus of knowledge claims emerging from a split half analysis is the extent to which items on the scale measure the same construct. Shogren, Thompson, Wehmeyer, Chapman, Tassé, and McLaughlin (2014) and Verdugo, Arias, Ibanez and Schalock, (2010) reported results from split-half reliability studies where SIS—A items were randomly divided into two sets, but the entire instrument was administered. Both research teams reported high correlations between scores on each half of the scale, thus providing additional evidence to support the reliability of the SIS—A. No split half reliability studies have been published using data from the SIS—C.

Test-retest reliability. Studies reporting findings from test-retest reliability and interrater reliability require comparing results from two administrations of an assessment. The correlations coefficients that are generated from an analysis of data from these studies provide an unambiguous measure of an assessment scales' consistency across people and time. Although test-retest results for the SIS—C are reported in the SIS—C user's manual (Thompson et al., 2016), no such results have been published in the professional literature. Results from test-retest reliability studies of the SIS—A, however, are available. Test-retest studies by Soltani, Kamali, Chabok, and Ashayeri (2013) and Verdugo et al. (2010) provided results for all of the subscale scores and the composite score (i.e., Support Needs Index Score). Both research teams reported that all coefficients fell within Cicchetti and Sparrow's (1981) good or excellent range. Cicchetti and Sparrow 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.

TABLE 1
Psychometric Properties of SIS—A and SIS—C Reported in Peer-Reviewed Journals (2002-2017)

Author (year)	Scale		Reliability				Content & Criterion Validity		Construct Validity			
	SIS—A	SIS—C	Internal Consist	Split Half	Test Retest	Inter-rater	Content	Criterion	Internal Structure	Convergent/Discriminate	Predict Validity	External Validity
1. Arnkelsson & Sigurdsson (2014)	•		•					•				•
2. Arnkelsson & Sigurdsson (2016)	•		•					•				•
3. Bossaert et al. (2009)	•		•				•	•	•	•		•
4. Brown et al. (2009)	•								•	•		
5. Chou et al. (2013)	•		•				•	•	•		•	•
6. Claes et al. (2009)	•					•			•	•		•
7. Claes et al. (2012)	•							•				•
8. Cruz et al. (2013)	•									•		•
9. Cruz et al. (2010)	•		•							•		•
10. Giné et al (2017)		•							•			
11. Giné et al. (2014)	•									•	•	•
12. Guillen et al. (2017)		•	•				•	•	•			•
13. Guillen et al. (2015)		•	•						•			•
14. Guscia et al. (2006)	•							•		•		

(continued)

Author (year)	Scale		Reliability				Content & Criterion Validity		Construct Validity			
	SIS—A	SIS—C	Internal Consist	Split Half	Test Retest	Inter-rater	Content	Criterion	Internal Structure	Convergent/Discriminate	Predict Validity	External Validity
15. Harries et al. (2005)	•								•	•		
16. Jenaro et al. (2011)	•		•			•			•	•	•	•
17. Kuppens et al. (2010)	•								•	•		•
18. Lamoureux-Hebert & Morin (2009)	•		•							•		•
19. Lamoureux-Hebert et al. (2010)	•									•		•
20. Lombardi et al. (2016)	•										•	•
21. Morin & Cobigo (2009)	•					•						
22. Seo, Shogren, Little et al. (2016)	•	•							•	•		
23. Seo, Wehmeyer et al. (2017)	•	•						•	•			
24. Seo, Shogren, Wehmeyer et al. (2016)	•	•						•	•			
25. Seo, Shogren et al. (2017)	•									•		
26. Shogren, Wehmeyer et al. (2017)		•	•					•	•	•		•
27. Shogren et al. (2015)		•							•	•		
28. Shogren et al. (2016)	•								•	•		

(continued)

Author (year)	Scale		Reliability				Content & Criterion Validity		Construct Validity			
	SIS—A	SIS—C	Internal Consist	Split Half	Test Retest	Inter-rater	Content	Criterion	Internal Structure	Convergent/Discriminate	Predict Validity	External Validity
29. Shogren, Shaw et al. (2017)		•							•	•	•	•
30. Shogren et al. (2014)	•		•			•			•	•	•	
31. Simões et al. (2016)	•									•	•	•
32. Smit et al. (2011)	•		•						•	•	•	•
33. Soltani et al. (2013)	•		•		•				•			•
34. Tassé & Wehmeyer (2010)	•									•		•
35. Thompson et al. (2002)	•		•				•	•	•	•		
36. Thompson et al. (2016)	•								•			
37. Thompson et al. (2008)	•					•						
38. Thompson et al. (2014)		•	•				•		•	•		
39. Tremblay & Morin (2015)	•										•	•
40. Verdugo, Arias et al. (2016)		•							•			•
41. Verdugo et al. (2010)	•		•		•	•	•	•		•		•
42. Verdugo, Guillen et al. (2016)		•							•			
43. Wehmeyer et al. (2009)	•					•				•	•	
44. Weiss et al. (2009)	•										•	•

Interrater reliability. Interrater reliability studies were slightly more common than test-retest studies, with results reported from six investigations. All of the interrater studies were focused on the SIS—A and none included data from the SIS—C. There are, however, extensive findings on SIS—C interrater reliability in the SIS—C user’s manual (Thompson et al., 2016).

In regard to SIS—A interrater reliability investigations, strong correlation coefficients for both the subscale scores and the overall score were reported in all of the studies. The majority of coefficients fell within Cicchetti and Sparrow’s (1981) excellent range (i.e., above .74), with a minority within the good range (i.e., .60 to .74), a small number of outliers in the fair range (i.e., .40 to .59), and none in the poor range (i.e., below .40). For example, in Thompson, Tassé, and McLaughlin’s (2008) study (discussed later in more detail), 42 coefficients were reported based on differing conditions under which data were collected and analyzed. In their study, 67% of the coefficients were in the excellent range, 24% were in the good range, and 9% were in the fair range.

Two interrater studies deserve special highlighting due to the uniqueness of their approach. Thompson et al. (2008) attempted to parcel out the variance that resulted from having different interviewers and different respondents involved in SIS—A interviews. They conducted some SIS—A interviews with the same interviewer but different respondents, others with the different interviewers and the same respondents, and still others with different interviewers and different respondents. Their data revealed the SIS—A proved to be highly reliable under each condition. This led them to posit that ensuring that interviewers are properly trained to administer the SIS—A may be the most critical influence on interrater reliability.

Claes, Van Hove, van Loon, Vandavelde, and Schalock (2009) used separate interviewers with staff as respondents and people with disabilities (who were the focus of the assessment) as respondents to investigate how respondent characteristics might influence SIS—A scores and inter-respondent reliability. They found that ratings from the two groups were quite reliable in the sense that the people whom the staff rated as having comparatively higher or lower support needs relative to others were the same people who rated themselves as having higher or lower support needs relative to others. However, people with disabilities consistently rated themselves as having less intense support needs than did their paid support staff.

It is important to remember that correlation coefficients reflect the linear relationship between two sets of scores and, therefore, correlational studies provide insight into the magnitude of proportionality between two sets of scores but do not provide information relevant to differences in numerical magnitude. It is unclear whether people with disabilities in Claes et al.’s (2009) study were inclined to underestimate their support needs (perhaps because of the stigma associated with needing assistance, or maybe they were not fully aware of their needs for support) or if the paid staff were inclined to overestimate the support people needed (perhaps due to a caregiver disposition of overprotectiveness). This study highlights the importance of obtaining perspectives from multiple respondents, and the importance of training interviewers to probe respondents when conflicting information surfaces.

Validity

Findings on the validity of the SIS—A and SIS—C reflect the extent to which the assessments are authentic reflections of the intensity of support people need to participate in age-appropriate, community-based settings and activities. Conclusions about the validity of any assessment can only be made by considering evidence from multiple sources. The stronger and more comprehensive the body of evidence, the more confident one can be in the validity of an assessment.

Content validity. As mentioned previously, content validity is evidenced by documenting that items on an assessment reflect the universe of potential items, and efforts to establish content validity must be undertaken when an assessment is being developed (i.e., when subscales are conceptualized and items are created). Two articles in this review documented such efforts. Two years before the SIS—A was first published, Thompson et al. (2002) described steps taken to develop the scale. They reported that a literature review of approximately 1,500 sources was undertaken to identify candidate items, dimensions on which to assess support needs, and components/categories of support. They also described the process and results of a Q sort involving 50 experts, where candidate items were evaluated and categorized, and potential content gaps were identified as well. Additionally, feedback from people with disabilities was solicited to ensure that their perspectives were taken into account. Finally, a pilot study was completed to get a sense of how users perceived the candidate items in terms of usefulness and redundancy.

In regard to efforts made to document the content validity of the SIS—C, 2 years prior to its publication, Thompson, Wehmeyer et al. (2014) reported that identifying items from the SIS—A that were applicable to children was their first step. Next, they completed the same literature review and Q-sort activities were undertaken for SIS—A to identify, sort, and refine additional candidate items. Finally, a pilot test to obtain the perspectives of prospective users (e.g., people with disabilities, service provider staff, jurisdictional decision makers) was completed.

There are also ways to evaluate content validity after an assessment scale has been published. Verdugo et al. (2010) employed a panel of judges to evaluate the relevance of items on the SIS—A Spanish Version to assure the content validity of the SIS—A transferred to the Spanish context. The most common approach to investigating content validity was through examining the relationship between each item and the other items on the assessment scale to determine whether each item is contributing something unique (and, therefore, is not redundant with other items). Bossaert, Kuppens, Buntinx, Molleman, Van den Abeele, and Maes (2009) and Chou, Lee, Chang, and Yu (2013) completed an item analysis for the Dutch and Complex Chinese versions of the SIS—A, respectively, and Guillen, Adam, Verdugo, and Gine (2017) and Thompson, Wehmeyer et al. (2014) did the same for the Spanish and English versions of the SIS—C, respectively. All four research teams reported that their data revealed that the items on the SIS scales had strong technical properties consistent with content validity.

Criterion-related validity. Within the methodological literature, one can find a variety of examples of indicators of criterion validity, all of which emphasize the importance of demonstrating a relationship between the measure of interest and

a benchmark measure. In some studies, multiple measures are used. For example, a new assessment of reading achievement might be positively correlated with established standardized assessments of reading achievement and recognized performance measures such as fluency in reading aloud. To determine what constitutes an appropriate criterion measure, one must consider the nature of construct being measured, as well as the body of research that has come before.

In regard to support needs assessment, 19 articles cited in Table 1 included findings comparing SIS scores to scores from adaptive behavior scales or other measures of personal competence. Several authors of these studies indicated that they were investigating criterion-related validity (i.e., evidence of a relationship to a benchmark criterion); others indicated they were investigating convergent validity (i.e., evidence that constructs that are supposed to be related are, in fact, related). To put it another way, some research teams considered their investigation into the relationships between measures of personal competence and support needs to fit under the criterion validity umbrella, and others classified the exact same type of investigation under the convergent validity umbrella.

Our purpose is not to sort out the confusion that exists in the methodological literature between the labels that are used to describe different types of validity. Our purpose, rather, is to summarize the collective findings from investigations of the validity of the SIS—A and SIS—C in a coherent manner. Therefore, we coded research studies the following way: (a) investigations of criterion-related validity were those where SIS—A or SIS—C scores were compared to a different measure of support needs, and (b) investigations of convergent validity were those where researchers used indicators of personal competence, most often scores from adaptive behavior scales or IQ tests, as their measure of comparison.

For purposes of criterion-related validity in the manner in which we are defining it, the SIS—A and SIS—C research teams shown in Table 1 used an array of support need measures for their criterion benchmarks. In regard to the SIS—A, Arnkelson and Sigurdsson (2014, 2016) used a seven-level measure of assistance needed that was associated with a legacy model for resource allocation in Iceland, Chou and colleagues (2013) used a measure of assistance required in relationship to a variety of Instrumental Activities of Daily Living (IADL), Guscia, Harries, Kirby, Nettelbeck, and Taplin (2006) compared SIS—A scores with scores from the Service Need Assessment Profile (SNAP; a different support needs assessment scale), and Thompson et al. (2002) and Verdugo et al. (2010) used Likert-type measures of perceived support needs. For the SIS—C, both Guillen et al. (2017) and Shogren, Wehmeyer et al. (2017) used the same Likert-type measures of perceived support needs that were first employed by Thompson et al. (2002). The results from these investigations consistently revealed moderate-sized correlations between the criterion measure of supports and the SIS—A and SIS—C scores, which provides collective evidence for the criterion-related validity of the scales. Correlation coefficients of moderate size are robust indicators of criterion validity, as high correlation coefficients might suggest that an assessment scale replicates (i.e., is redundant with) a criterion measure, and low correlation coefficients would suggest a weak association between the assessment scale and the construct on which the criterion measures are based.

Construct validity: Internal structure. Support for an assessment scale's internal structure come from research findings that demonstrate a strong relationship between the construct being measured and an assessment scale's test items and the subscale scores. Statistical analyses revealing that the items share variance in ways that match the defined construct reflect positively on an assessment scale's internal structure. There were 25 articles reporting findings relevant to assessing the internal structure of the SIS—A and SIS—C. Specifically, 13 research teams collected data from the SIS—A, nine collected data from the SIS—C, and three collected data on both the SIS—A and SIS—C.

Eight SIS—A and SIS—C research teams investigated the evidence supporting the internal structure of the scales by examining the magnitude of subscale correlations. When subscales are included in an assessment's structure, there should be evidence that items within each subscale share variance with one another that they do not share with items on the other subscales. The range of coefficients reported (.45 to .87) on the SIS—A in the very first study by Thompson et al. (2002) has largely been replicated by every other research team investigating both the SIS—A and SIS—C. Brown, Ouellette-Kuntz, Bielska, and Elliot (2009) reported somewhat higher correlation coefficients among the SIS—A subscales than other research teams (.72 to .88), but their sample ($n = 40$) was smaller than most. Even with the higher correlations, their highest coefficients did not reach the threshold (i.e., .90) that MacEachron (1982) cautioned might suggest two subscales lacked independence. The range (.44 to .82) reported by Smit, Sabbe, and Prinzie (2011) was interesting because their numbers practically mirrored those reported for the original standardization sample (Thompson et al., 2004), despite the fact that they collected data on a translated version of the SIS—A and on a sample of people with physical disabilities whose characteristics were significantly different than those in Thompson et al.'s standardization sample.

To summarize, seven research teams investigating the relationship among subscales scores on the SIS—A reported moderately high (i.e., .40 to .90) positive correlation coefficients. These findings held true as well in the one study by Thompson, Wehmeyer et al. (2014) where correlation coefficients among the subscales of the SIS—C were reported. Although the correlations between subscales on the SIS—A and SIS—C are relatively high, they are not so high as to suggest that the subscales lack independence from one another. Put another way, the internal structures of the SIS—A and SIS—C are supported by investigations of subscale correlation because there is evidence that the subscales (a) are measuring the same overall construct, but at the same time (b) are sufficiently independent to suggest that different aspects of the broader construct are being measured by each subscale.

Other research teams have used latent modeling techniques to investigate the internal structure, specifically the factor structure of the SIS—A and SIS—C. The collective findings from these 15 investigations have suggested that both scales have a robust internal structure. Harries, Guscia, Kirby, Nettelbeck, and Taplin (2005) were the only research team to conclude that their data supported a unidimensional factor structure. They combined data from SIS—A scores and two adaptive behavior scales, and they conducted an exploratory factor analysis (EFA) to investigate the factor structure of their combined data. It is important to

acknowledge, however, that combining data from three scales could mask a factor structure that might have emerged if data from only one scale had been entered.

Bossaert et al. (2009) also used EFA in their analysis of the SIS—A data from a diverse sample in terms of disability characteristics, and their data supported a four-factor solution. Both Harries et al. and Bossaert et al.'s findings must be considered in light of the limitations of EFA. Because EFA does not place any constraints on data, EFA can be an arbitrary approach to obtaining an understanding of a construct. Experts in latent modeling methods stress that confirmatory factor analysis (CFA), where factor structures are tested that are based on prior research or theory, provide a more meaningful approach to investigating the factor structure of a scale (and a construct). The findings from studies using CFA have confirmed the factor structure (i.e., subscale structure) on which the SIS—A and SIS—C were established.

The findings from five research teams reporting on the factor structure of the SIS—A that have used CFA with large sample sizes have supported the six-factor structure that corresponds to the six SIS—A subscales (Home Living, Community Living, Lifelong Learning, Employment, Health and Safety, and Social Activities). When the six-factor solution was tested alongside alternative structures (e.g., unidimensional, four-factor), a variety of goodness of fit statistics indicated the superiority of the six-factor solution (e.g., see Kuppens, Bossaert, Buntinx, Molleman, Van den Abbeele, & Maes, 2010). The only exception was Shogren, Seo, Wehmeyer, Thompson, and Little's (2016) study, where data from the Supplemental Protection and Advocacy (P&A) Scale was included in the analysis with data from the six SIS—A subscales. In that study, the six subscale-related factors remained intact, but a seventh factor corresponding with the Supplemental P&A Scale emerged as well. This finding, along with other evidence, prompted Shogren and colleagues to recommend that strong consideration be given to including the Supplemental P&A Scale in the standardized portion of the SIS—A when the scale is revised.

The 10 studies that have reported findings relevant to the factor structure of the SIS—C all produced results supportive of the seven-factor structure that aligns with the SIS—C's seven subscales. It is important to note that several of these studies were not designed to test the factor structure explicitly, but, rather, used latent variable modeling methods to investigate other research questions. These methods, however, were contingent on evidence of measurement invariance related to the factor structure. Therefore, these studies provide additional evidence supporting the internal structure of the SIS—C.

The most direct and comprehensive investigation into the factor structure of the SIS—C was completed by Verdugo, Guillen, Arias, Vicente, and Badia (2016). They found that the seven-factor solution that was aligned with the seven subscales was the best of several competing models. Moreover, a correlated first-order factor model (where the seven factors were correlated with one another) was superior to a hierarchical factor model involving a global support needs factor (i.e., where the seven factors were subordinate to an upper-level factor).

Although there is room for additional investigation into the internal structure of both the SIS—A and SIS—C, the cumulative evidence from the subscale correlation studies and the factor structure studies suggest that the internal structures of both instruments have

been empirically validated. Findings across multiple studies reveal a consistent pattern of extensive shared variance among items and subscales, and this provides evidence that different aspects of a common construct are being measured by the SIS—A and SIS—C.

Convergent and discriminate validity. Convergent and discriminate validity go hand in hand. Convergent validity is established by collecting data that show that constructs that theoretically should be related are, in fact, related. Discriminate validity is established by collecting data showing that constructs that theoretically should have no relationship with one another, in fact, have low correlation coefficients. When both convergent and discriminate validity are established, a strong case can be made that an assessment tool possesses construct validity (i.e., it measures what it purports to measure).

Convergent validity has been evaluated most often by comparing SIS—A and SIS—C scores with scores on adaptive behavior (AB) scales and/or IQ tests. It is hypothesized that there should be a negative correlation between support needs and measures of personal competency, meaning that people with relatively more intense support needs (higher SIS scores) show relatively less competence in regard to adaptive behavior (lower AB scores) or intelligence (lower IQ scores). Investigations comparing IQ scores to the SIS—A (Chou et al., 2013; Lamoureux-Hebert, & Morin, 2009; Thompson et al., 2002) or to the SIS—C (Thompson et al., 2014) have used IQ ranges (e.g., 1 = Mild; 2 = Moderate; 3 = Severe; 4 = Profound) as the unit of analysis, not actual scores. The results of these investigations have all revealed the expected relationship—namely, that IQ has a moderate to high negative correlation with SIS scores.

Fourteen (14) AB studies have been conducted with the adult version of the SIS, but Thompson et al.'s (2014) investigation was the only one comparing AB scores with assessment results from the children's version. An array of scales measuring adaptive behavior skills were used across the studies, including the Scales of Independent Behavior Revised or SIB-R (e.g., see Brown et al., 2009), the Vineland Adaptive Behavior Scales (e.g., see Claes et al., 2009), the Inventory of Client and Agency Planning (ICAP; e.g., see Giné, Font, Guardia-Olmos, Balcells-Balcells, Valls, & Carbo-Carrete, 2014), and the Adaptive Behavior Scale-Residential/Community (ABS-RC; e.g., see Simões, Santos, Biscaia, & Thompson, 2016). In all studies, the anticipated negative correlation coefficients were generated (i.e., greater adaptive skills = less intensive support needs).

Although the magnitude of correlation coefficients varied somewhat from study to study and was partially dependent on which measure of adaptive behavior was used and whether total scores or subscale scores on the instruments were compared, collectively, the findings were consistent overall. For example, when reaching their conclusions, Brown et al. (2009) and Harries et al. (2005) specifically stated that their findings were aligned with findings presented in Thompson et al.'s (2004) SIS—A user's manual (i.e., findings based on the original standardization samples). Despite the consistent findings among the studies, two of the of research teams reached different conclusions about their findings than the other research teams. Brown et al. (2009) and Harries et al. (2005) concluded that the correlations between SIS—A scores and AB scores were sufficiently high to suggest that the scales were not measuring different constructs. The other 13 research teams reached the opposite conclusion (e.g., see Kuppens et al., 2010; Lamoureux-Hebert, Morin, & Crocker, 2010; Verdugo et al., 2010).

Differences in conclusions about whether the SIS scales are measuring a different construct than what is measured by adaptive scales appears to hinge on interpretations of what constitutes a large, moderate, or small correlation. Cohen (1988) provided guidelines that are often cited, although he advised that the magnitude of correlation coefficients should always be considered within the context of the variables that are being investigated. Cohen indicated that 0.5 or above is large, 0.5 – 0.3 is moderate, 0.3 – 0.1 is small, and anything smaller than 0.1 is insignificant. The correlations of SIS scores with AB scores, whether comparing subscale or total scores, have consistently been shown to be in the high end of Cohen’s moderate range, or the low end of Cohen’s high range. For example, Verdugo et al. (2010) reported the correlation coefficients obtained for SIS and ICAP scores ranged from $-.498$ and $-.589$, and Lamoureux-Hebert et al. (2010) reported that correlations between the SIS—A total score and SIB-R subscale scores ranged from $-.18$ to $-.36$. Perhaps Simoes et al. (2016) summarized the literature and the conclusions around this question best when stating,

The relationship between support needs and adaptive behavior has been investigated by multiple researchers. There is consensus that the two are related, but different, constructs. ... Generally speaking, people with greater skills will have less intense support needs, and those with lesser skills will have more intense support needs. There are, however, many influences on support needs other than the degree of adaptive skill acquisition. (p. 850)

Discriminate validity has been primarily investigated by examining the relationship of age and gender to SIS—A and SIS—C scores (e.g., Kuppens et al., 2010; Shogren, Wehmeyer et al., 2017; Thompson et al., 2002). All three research teams reporting findings on gender confirmed that research evidence supported discriminate validity. Gender has been shown to have insignificant correlations with overall SIS—A and SIS—C scores, although a more nuanced analysis may eventually reveal that certain items are related to gender. For example, Shapiro (2018) reported that the U.S. Justice Department recently analyzed data indicating that females may be at greater risk for both physical and sexual abuse and, therefore, would be expected to have more intense support needs than males in regard to supports needed to *protect self from exploitation* (an item from both the SIS—A and SIS—C).

Findings in regard to the relationship between age and SIS—A and SIS—C scores have been reported by nine research teams. Age been shown to have a very low correlation with SIS—A scores, and this supports discriminate validity because people with disabilities with relative high intensity and relatively low intensity support needs are distributed across age groups in adulthood. However, the relationship between age and support needs in adulthood needs more careful study, as health issues should impact support needs as people move into older adulthood.

In contrast to the SIS—A, a correlation between SIS—C scores and with age provides an indicator of *convergent validity* because younger children generally require more supervision than older children regardless of disability status. This was the rationale for creating separate norms for different age groups on the SIS—C. Giné et al. (2017) and Shogren et al. (2015) have reported comprehensive analyses of the relationship between the age of children and intensity of support needs as measured

by the SIS—C. Their data clearly supports the convergent validity of the SIS—C because age was correlated with SIS—C scores in both studies. However, findings from these studies also suggest that the cultural context may mediate the influence of age to a certain extent. Both studies provided evidence that the same set of SIS—C items could be used to measure support needs of children in the United States and Catalonian region of Spain, but there were age-related differences in support needs in the two samples. Data from the U.S. sample supported separate norms for six different age cohorts (5-6, 7-8, 9-10, 11-12, 13-14, & 15-16), whereas data from the Catalonia sample supported only two sets of age norms (5-10 & 11-16).

In summary, research findings have supported both the convergent and discriminate validity of the SIS—A and SIS—C. In terms of convergent validity, research findings show that SIS—A and SIS—C scores have a negative correlation with measures of personal competence. Due to the volume of studies examining the relationship between SIS scores and adaptive behavior, there is now a strong empirical basis on which to conclude that adaptive behavior and support needs are related, but different, constructs. Both the SIS—A and SIS—C show good discriminate validity properties by having little relationship to gender. In terms of the relationship to age, the very low correlations between SIS—A scores and age provide evidence of discriminate validity, whereas the negative correlation between SIS—C scores and age provide evidence of convergent validity (due to the differences in the support needs of children and adults).

Predictive validity. Predictive validity, like criterion, convergent, and discriminate validity, is also evidenced by examining relationships between the measure of interest (i.e., the SIS—A and/or SIS—C in this case) and other measures. For the purpose of this review we identified predictive studies as those that investigated the capacity of the SIS—A or SIS—C to differentiate between groups of people that were known to differ on support needs. There were 11 studies that aligned with our definition of predictive validity, two of which were focused on the relationship between SIS—A scores and quality of life (QOL) indicators, three that investigated how SIS—A scores related to funding levels, and six that investigated the relationship between SIS—A and SIS—C scores with diagnostic categories and disability characteristics.

The two studies exploring how SIS scores predicted QOL used the Personal Outcomes Scale (POS; van Loon, Van Hove, Schalock, and Claes, 2015) as the dependent variable. The POS is a comprehensive QOL measure that provides scores under two conditions: self-report (based on self-appraisal) and direct observation (based on life events and circumstances). Lombardi, Croce, Claes, Vandeveld, and Schalock (2016) and Simões et al. (2016) used regression analysis to investigate the predictive strength of SIS—A scores and other variables in relation to POS scores (i.e., QOL outcomes). Both research teams found that greater support needs were associated with poorer QOL outcomes. The research teams did not conclude that more intense support needs caused a poorer quality of life, but, rather, it is the *absence of supports to meet those needs* that results in diminished life experiences and opportunities. Put another way, these data suggest people with greater support needs are more likely to experience a greater magnitude of unmet support needs, which results in diminished QOL outcomes.

Three research teams examined the SIS—A's relationship to the allocation of public funds. Giné et al (2014) analyzed data collected in Catalonia to investigate funding approaches using information from the ICAP and the SIS—A. They concluded that the SIS—A provided a more sensitive measure to identifying differing levels of need. Based on results from their cluster analysis, they proposed a funding structure to better align resource allocation with extent of need. Working in the Taiwanese context, Chou et al. (2013) demonstrated how results from the SIS—A could be applied to create a more equitable and defensible approach to distributing public funding compared to the legacy approach that was currently being used. Wehmeyer and colleagues (2009) reached a similar conclusion in the United States. By applying correlational and regression analysis to jurisdictional level data and SIS—A assessment data, they found that data from the SIS—A explained more of the variance related to extraordinary support needs than the adapted version of the Developmental Disability Profile (DDP) that the jurisdiction currently used to inform disbursement of public funds.

The final collection of studies related to predictive validity of the SIS—A and SIS—C used a variety of diagnostic-based categories as the dependent variable. Jenaro, Cruz, Perez, Flores, and Vega (2011) found that the SIS—A was a good predictor of a “length of disease” variable for people with mental illness. The longer people in the sample had been diagnosed and receiving treatment, the higher their intensity of support needs. Smit et al. (2011) found that the SIS—A was a strong predictor of number of disability diagnoses, with those with more diagnoses having more intense needs for support. Shogren et al. (2014) and Tremblay and Morin (2015) analyzed SIS—A data along with data from measures that were being used to identify levels of specialized care. Both research teams found that those with more intense support needs had higher care needs. Weiss, Lunskey, Tassé, and Durbin (2009) investigated whether SIS—A scores predicted clinician ratings of overall level of need. They reported that,

of all the standard SIS scores, Home Living Activities was the strongest predictor of clinician-ranked level of need. Individuals who have high scores on this subscale need more intense supports in self-care and home-living activities, such as dressing, hygiene, using appliances and equipment, and preparing and eating food. (p. 939)

Finally, Shogren, Shaw et al. (2017) investigated the degree to which latent clusters were present in SIS—C data that were collected on a sample of children with a dual diagnosis of intellectual disability and autism spectrum disorder (ASD). They found support for four clusters based on relative intensity of support need, and they contrasted their findings with the three-level classification categories identified for ASD in the most recent edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*; American Psychiatric Association, 2013).

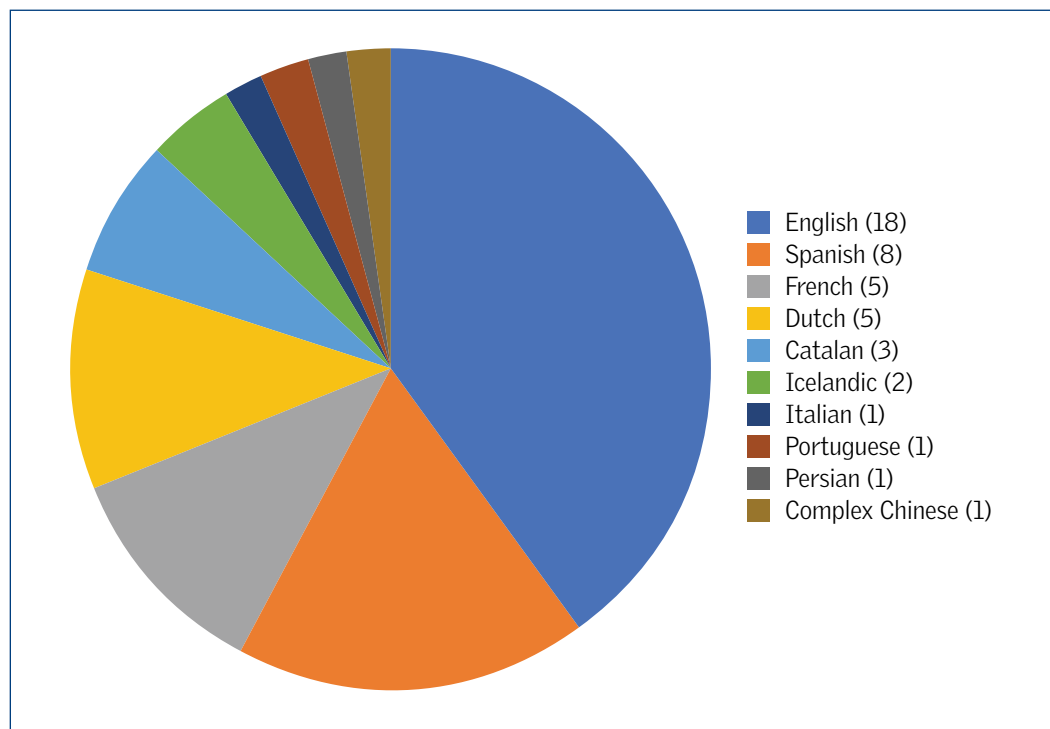
When considering the 11 studies where the predictive validity of the SIS—A and SIS—C were investigated, it is striking that none of the studies were structured in a way to investigate the accuracy of prediction for a measure collected at a future point in time. Despite the fact that the evidence base could be expanded, the collective findings from these studies confirm the predictive validity of the

SIS—A and SIS—C. The SIS scales provide a reasonable way to differentiate groups of people who are known to differ on the intensity of support they need.

External validity. Evidence for external validity is focused on the extent to which an assessment scale can be applied with people and in settings other than for whom it was originally developed. The SIS—A and SIS—C were originally written in English and the standardization samples consisted of children and adults from North America with a primary disability of intellectual disability. Figure 3 shows the number of articles published in peer-reviewed journals that are (a) based on data from the SIS—A and SIS—C assessments administered in 11 different languages with (b) findings relevant to the psychometric properties of the SIS scales. The total is 45, one more than the number of articles listed in Table 1, because Guillen and her colleagues (2017) compared data from Spanish and Catalan versions of the SIS—C. Therefore, their article was included in both the Spanish and Catalan tallies.

The external validity of the SIS scales is also evidenced by the array of different disability conditions that have been reported in the samples on which data were collected. Although no effort was made to catalogue all of the diagnoses reported across the 44 studies, several studies provided specific information on primary and secondary diagnoses (e.g., Arnkelsson & Sigurdsson, 2016; Kuppens et al., 2010; Lombardi et al., 2016; Thompson, Wehmeyer et al., 2014). In no investigation did the researchers actually verify any diagnosis through data collection. Rather, diagnostic information was obtained through either a record review or a report from a respondent.

FIGURE 3
Number of articles published using data from the English and translated versions of the SIS—A and SIS—C.



All of the studies that specifically set out to investigate how the SIS—A or SIS—C functioned with people with disabilities other than intellectual disability generated findings that indicated the SIS scales retained strong psychometric properties with the new populations. People with mental illness were the focus of investigations by Arnelsson and Sigurdsson (2014); Cruz, Jenaro, Perez, and Robiana (2010); Cruz, Perez, Jenaro, Flores, and Vega (2013); and Jenaro et al. (2011). People with physical-motor disabilities were participants in studies by Arnelsson and Sigurdsson (2016) and Smit et al. (2011). Bossaert et al. (2009) did not set out to investigate how well the SIS—A worked with people with acquired brain injury, but nearly half (47.81%) of their sample reported this diagnosis. Finally, Shogren, Wehmeyer, et al. (2017) and Shogren, Shaw, et al. (2017) had children with ASD as the focus of their studies.

The fact that strong psychometric indicators are evident from investigations using translated versions of the SIS—A and SIS—C provides evidence that the scales have proven to be applicable across a variety of languages, countries, and cultures. It is also a testament to the “committee approach” to translation procedures that have been used by researchers when translating the SIS. This approach to translation was originally proposed by Tassé and Craig (1999) and later refined by Tassé and Thompson (2010). Certainly, more work needs to be done to validate the SIS scales with populations other than children and adults with intellectual disability. However, the evidence thus far suggests that the utility of the SIS scales may transcend disability boundaries. At this time, there is every reason to be confident that the SIS scales will continue to be found to be applicable to measuring and understanding the support needs of people from multiple and diverse cultural and disability populations.

Conclusion

The SIS—A and SIS—C are the only standardized assessment instruments in the field of intellectual disability and developmental disabilities that measure the intensity of a person's support needs. The SIS scales are considered to be standardized assessments because (a) all people who are assessed must be assessed on the same set of items; (b) a uniform procedure, in the form of a semistructured interview, must be used to collect data; and (c) standard scores are generated that allow comparisons of a person's support needs to that of a representative sample of people with IDD.

Forty-four articles relevant to the psychometric properties of the SIS scales have been published in refereed, peer-reviewed journals, where the details of research procedures and subsequent findings have been communicated. Multiple researchers have authored these studies, and data have been collected in wide variety of geographical and cultural settings. As was alluded to earlier, construct validity of any assessment tool is developed over time through an accumulation of multiple sources evidence. Although there is still much to learn, there is convincing evidence that the SIS scales provide a reliable and valid means to measure the intensity of the support needs of children and adults with intellectual disability and related developmental disabilities.

References

(References for citations reporting psychometric findings are also listed in the Appendix.)

- Aiken, L. R., & Groth-Marnat, G. (2005). *Psychological testing and assessment* (12th ed.). Needham Heights, MA: Allyn & Bacon.
- Alonso, M., Gomez, L., & Navas, P. (2012). The Supports Intensity Scale: Lessons from Spain. *Journal of Intellectual Disability Research*, *56*, 799–799.
- American Association on Intellectual and Developmental Disabilities. (2017a). *International SIS use: International users*. Retrieved from https://aaid.org/sis/international#.WmZV85M-e_p
- American Association on Intellectual and Developmental Disabilities. (2017b). *States/Provinces using SIS in North America*. Retrieved from https://aaid.org/sis/sisonline/states-using-sis#.WmZW5JM-e_p
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.
- Arnkellsson, G., & Sigurdsson, T. (2014). The validity of the Supports Intensity Scale for adults with psychiatric disabilities. *Research in Developmental Disabilities*, *35*, 3665–3671. <http://doi.org/10.1016/j.ridd.2014.09.006>
- Arnkellsson, G., & Sigurdsson, T. (2016). The validity of the Supports Intensity Scale for adults with motor disability. *American Journal of Intellectual and Developmental Disabilities*, *121*(2), 139–150. <http://dx.doi.org/10.1352/1944-7558-121.2.139>
- Brown, H. K., Ouellette-Kunz, H., Bielska, I., & Elliott, D. (2009). Choosing a measure of support need: Implications for research and policy. *Journal of Intellectual Disability Research*, *53*(11), 949–954. <http://dx.doi.org/10.1111/j.13652788.2009.01216.x>
- Chou, Y. C., Lee, Y. C., Chang, S. C., Yu, A. P. (2013). Evaluating the Supports Intensity Scale as a potential assessment instrument for resource allocation for persons with intellectual disability. *Research in Developmental Disabilities: A Multidisciplinary Journal*, *34*(6), 2056–2063. <http://dx.doi.org/10.1016/j.ridd.2013.03.013>
- 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*(2), 127–137.
- Claes, C., Van Hove, G., Van Loon, J., Vandeveld, S., & Schalock, R. L. (2009). Evaluating the inter-respondent (consumer vs. staff) reliability and construct validity (SIS vs. Vineland) of the supports intensity scale on a Dutch sample. *Journal of Intellectual Disability Research*, *53*, 329–338. <http://dx.doi.org/10.1111/j.1365-2788.2008.01149.x>

- Clark, G. M., & Kolstoe, O. P. (1995). *Career development and transition education for adolescents with disabilities* (2nd ed.). MA: Allyn and Bacon.
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. New Jersey, NJ: Lawrence Erlbaum Associates Publishers. Retrieved from <http://imaging.mrc-cbu.cam.ac.uk/statswiki/FAQ/effectSize>.
- Cruz, M., Perez, M. D., Jenaro, C., Flores, N., & Vega, V. (2013). Identification of the support needs of individuals with severe mental illness using the supports intensity scale. *Revista latino-americana de enfermagem*, *21*(5), 1137–1143. <http://dx.doi.org/10.1590/S0104-11692013000500017>
- Cruz, M., Jenaro, C., Perez, M. D., & Robaina, N. F. (2010). Applicability of the Spanish Version of the Supports Intensity Scale (SIS) in the Mexican population with severe mental illness. *Revista latino-americana de enfermagem*, *18*, 975–982.
- Davison, H. (2005). The Supports Intensity Scale. *Journal of Intellectual Disability Research*, *49*, 636–636.
- Fredericks, D. W., & Williams, W. I. (1998). New definition of mental retardation for the American Association of Mental Retardation. *Image—The Journal of Nursing Scholarship*, *30*(1), 53–56. <http://dx.doi.org/10.1111/j.1547-5069.1998.tb01236.x>
- Giné, C., Adam, A. L., Font, J., Salvador-Bertran, F., Baques, N., Oliveira, C., . . . Thompson, J. R. (2017). Examining measurement invariance and differences in age cohorts on the Supports Intensity Scale—Children’s Version—Catalan Translation. *American Journal on Intellectual and Developmental Disabilities*, *122*, 511–524. <http://dx.doi.org/10.1352/19447558122.6.511>
- Giné, C., Font, J., Guardia-Olmos, J., Balcells-Balcells, A., Valls, J., & Carbo-Carrete, M. (2014). Using the SIS to better align the funding of residential services to assessed support needs. *Research in Developmental Disabilities*, *35*(5), 1144–1151. <http://dx.doi.org/10.1016/j.ridd.2014.01.028>
- Guillen, V. M., Adam, A. L., Verdugo, M. A., & Giné, C. (2017). Comparisons between the Spanish and Catalan versions of the Supports Intensity Scale for Children (SIS—C). *Psicothema*, *29*, 126–132. <http://dx.doi.org/10.7334/psicothema2016.200>
- Guscia, R., Harries, J., Kirby, N., Nettelbeck, T., & Taplin, J. (2006). Construct and criterion validities of the service need assessment profile (SNAP): A measure of support for people with disabilities. *Journal of Intellectual and Developmental Disability*, *31*, 148–155. <http://dx.doi.org/10.1080/13668250600876459>
- Harries, J., Guscia, R., Kirby, N., Nettelbeck, T., & Taplin, J. (2005). Support needs and adaptive behaviors. *American Journal on Mental Retardation*, *110*(5), 393–404. [http://dx.doi.org/10.1352/0895-8017\(2005\)110%5B393:SNAAB%5D2.0.CO;2](http://dx.doi.org/10.1352/0895-8017(2005)110%5B393:SNAAB%5D2.0.CO;2)
- Jenaro, C., Cruz, M., Perez, M. D., Flores, N. E., & Vega, V. (2011). Utilization of the Supports Intensity Scale with psychiatric populations: Psychometric properties and utility for service delivery planning. *Archives of Psychiatric Nursing*, *25*(5), e9–e17. <http://dx.doi.org/10.1016/j.apnu.2011.05.002>

- Kuppens, S., Bossaert, G., Buntinx, W., Molleman, C., Van den Abbeele, A., & Maes, B. (2010). Factorial validity of the Supports Intensity Scale (SIS). *American Journal on Intellectual and Developmental Disabilities, 115*(4), 327–339. <http://dx.doi.org/10.1352/1944-7558-115.4.327>
- Lamoureux-Hebert, M., & Morin, D. (2009). Translation and cultural adaptation of the Supports Intensity Scale in French. *American Journal on Intellectual and Developmental Disabilities, 114*(1), 61–66. <http://dx.doi.org/10.1352/2009.114:61-66>
- Lamoureux-Hebert, M., Morin, D., & Crocker, A. (2010). Support needs of individuals with mild and moderate intellectual disabilities and challenging behaviors. *Journal of Mental Health Research in Intellectual Disabilities, 3*, 67–84. <http://dx.doi.org/10.1080/19315861003650558>
- Lombardi, M., Croce, L., Claes, C., Vandavelde, S., & Schalock, R. L. (2016). Factors predicting quality of life for people with intellectual disability: Results from the ANFFAS study in Italy [Special issue]. *Journal of Intellectual and Developmental Disabilities, 338–347*. <http://dx.doi.org/10.3109/13668250.2016.1223281>
- Luckasson, R., Borthwick-Duffy, S. A., 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). (2002). *Mental retardation: Definition, classification, and systems of supports* (10th ed.). Washington, DC: American Association on Mental Retardation.
- Luckasson, R., Coulter, D. L., Polloway, E. A., Reese, S., Schalock, R. L., Snell, M. E., . . . Stark, J. A. (1992). *Mental retardation: Definition, classification, and systems of supports* (9th ed.). Washington, DC: American Association on Mental Retardation.
- MacEachron, A. E. (1982). *Basic statistics in the human services*. Austin, TX: Pro-Ed.
- Rogers, A. C. (Secy.). (1910). Report of committee on classification of feeble-minded. *Journal of Psychoasthenics, 15*, 61–67.
- Salvia, J., Ysseldyke, J. E., & Bolt. S. (2010). *Assessment in special and inclusive education* (11th ed.). Boston, MA: Wadsworth/Cengage Publications.
- Schalock, R. L., Borthwick-Duffy, S. A., Bradley, V. J., Buntinx, W. H. E., Coulter, D. L., Craig, E. M., Gomez, S. C., Lachapelle, Y., Luckasson, R., Reeve, A., Shogren, K. A., Snell, M. E., Spreat, S., Tassé, M. J., Thompson, J. R., Verdugo-Alonso, M. A., Wehmeyer, M. L., & Yeager, M. H. (2010). *Intellectual disability: Definition, classification, and systems of supports* (11th ed.). Washington, DC: American Association on Intellectual and Developmental Disabilities.
- Seo, H., Little, T. D., Shogren, K. A., & Lange, K. M. (2016). On the benefits of latent variable modeling for norming scales: The case of the Supports Intensity Scale—Children’s Version. *International Journal of Behavioral Development, 40*, 373–384.
- Shapiro, J. (2018). *The sexual assault epidemic no one talks about*. Washington, DC: National Public Radio. Retrieved from <https://www.npr.org/2018/01/08/570224090/the-sexual-assault-epidemic-no-one-talks-about>

- Shogren, K. A., Seo, H., Wehmeyer, M. L., Palmer, S. B., Thompson, J. R., Hughes, C., & Little, T. D. (2015). Support needs of children with intellectual and developmental disabilities: Age-related implications for assessment. *Psychology in the Schools, 52*, 874–891. <http://dx.doi.org/10.1002/pits.21863>
- Shogren, K. A., Seo, H. S., Wehmeyer, M. L., Thompson, J. R., & Little, T. D. (2016). Impact of the protection and advocacy subscale on the factorial validity of the Supports Intensity Scale—Adult Version. *American Journal on Intellectual and Developmental Disabilities, 121*, 48–64. <http://dx.doi.org/10.1352/1944-7558-121.1.48>
- Shogren, K. A., Shaw, L. A., Wehmeyer, M. L., Thompson, J. R., Lang, K. M., Tassé, M. J., & Schalock, R. L. (2017). The support needs of children with intellectual disability and autism: Implications for supports planning and subgroup classification. *Journal of Autism and Developmental Disorders, 47*, 865–877. <http://dx.doi.org/10.1007/s10803-016-2995-y>
- Shogren, K. A., Thompson, J. R., Wehmeyer, M., Chapman, T., Tassé, M. J., & McLaughlin, C. A. (2014). Reliability and validity of the Supplemental Protection and Advocacy Scale of the Supports Intensity Scale. *Inclusion, 2*, 100–109. <http://dx.doi.org/10.1352/2326-6988-2.2.125>
- Shogren, K. A., Wehmeyer, M. L., Seo, H., Thompson, J. R., Schalock, R. L., Hughes, C., . . . Palmer, S. B. (2017). Examining the reliability and validity of the Supports Intensity Scale—Children’s version in children with autism and intellectual disability. *Focus on Autism and Other Developmental Disabilities, 32*, 293–304. <http://dx.doi.org/10.1177/1088357615625060>
- Simões, C., Santos, S., Biscaia, R., & Thompson, J. R. (2016). Understanding the relationship between quality of life, adaptive behavior, and support needs. *Journal of Developmental and Physical Disabilities, 28*, 849–870. <http://dx.doi.org/10.1007/s10882-016-9514-0>
- Smit, W., Sabbe, B., & Prinzie, P. (2011). Reliability and validity of the Supports Intensity Scale (SIS) measured in adults with physical disabilities. *Journal of Developmental & Physical Disabilities, 23*, 277–287. <http://dx.doi.org/10.1007/s10882-011-9227-3>
- Sobsey, D. (1987). *Ecological inventory exemplars*. Edmonton, Canada: University of Alberta.
- Soltani, S., Kamali, M., Chabok, A., & Ashayeri, H. (2013). Evaluating the validity and reliability of Persian Version of Supports Intensity Scale in adults with intellectual disability. *Journal of Kermanshah University of Medical Sciences, 17*(9), 555–562.
- Tassé, M. J., & Craig, E. M. (1999). Critical issues in the cross-cultural assessment of adaptive behavior. In R. L. Schalock (Ed.), *Adaptive behavior and its measurement: Implications for the field of mental retardation*. Washington, DC: American Association on Mental Retardation.

- Tassé, M. J., & Thompson, J. R. (2010). *Supports Intensity Scale for Children translation guidelines*. Paper presented at the 134th meeting of the American Association on Intellectual and Developmental Disabilities, Providence, RI.
- Thompson, J. R., Bradley, V., Buntinx, W. H. E., Schalock, R. L., Shogren, K. A., Snell, M. E., . . . Yeager, M. H. (2009). Conceptualizing supports and the support needs of people with intellectual disability. *Intellectual and Developmental Disabilities, 47*(2), 135–146. <http://dx.doi.org/10.1352/1934-9556-47.2.135>
- Thompson, J. R., Bryant, B., Campbell, E. M., Craig, E. M., Hughes, C., Rotholz, D. A., Schalock, R. L., Silverman, W., Tassé, M., & Wehmeyer, M. L. (2004). *The Supports Intensity Scale (SIS): User's manual*. Washington, DC: American Association on Mental Retardation.
- Thompson, J. R., Bryant, B., Schalock, R. L., Shogren, K. A., Tassé, M. J., Wehmeyer, M. L., . . . Rotholz, D. A. (2015). *Supports Intensity Scale—Adult Version: User's manual*. Washington, DC: American Association on Intellectual and Developmental Disabilities.
- Thompson, J. R., Doepke, K., Holmes, A., Pratt, C., Myles, B. S., Shogren, K. A., & Wehmeyer, M. L. (2017). *Person-centered planning with the Supports Intensity Scale—Adult version: A guide for planning teams*. Washington, DC: American Association on Intellectual and Developmental Disabilities.
- Thompson, J. R., Hughes, C., Schalock, R. L., Silverman, W., Tassé, M. J., Bryant, B., . . . Campbell, E. M. (2002). Integrating supports in assessment and planning. *Mental Retardation, 40*(5), 390–405. [http://dx.doi.org/10.1352/0047-6765\(2002\)040%3C0390:ISIAAP%3E2.0.CO;2](http://dx.doi.org/10.1352/0047-6765(2002)040%3C0390:ISIAAP%3E2.0.CO;2)
- Thompson, J. R., Schalock, R. L., Agosta, J., Teninty, L., & Fortune, J. (2014). How the supports paradigm is transforming the developmental disabilities service system. *Inclusion, 2*, 86–99. <http://dx.doi.org/10.1352/2326-6988-2.2.86>
- Thompson, J. R., Shogren, K. A., Schalock, R. L., Tassé, M. J., & Wehmeyer, M. L. (2017). *SIS—A annual review protocol*. Washington, DC: American Association on Intellectual and Developmental Disabilities.
- Thompson, J. R., Shogren, K. A., & Wehmeyer, M. L. (2017). Supports and support needs in strengths-based models of intellectual disability. In M. L. Wehmeyer & K. A. Shogren (Eds.), *Handbook of research-based practices for educating students with intellectual disability* (pp. 31–49). New York, NY: Routledge.
- Thompson, J. R., Shogren, K. A., Wehmeyer, M. L., Schalock, R. L., & Tassé, M. J. (in press). *SIS—A annual review protocol: A facilitator's guide*. Washington, DC: American Association on Intellectual and Developmental Disabilities.
- 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.

- Thompson, J. R., Wehmeyer, M. L., Hughes, C., Shogren, K. A., Little, T. D., Copeland, S. R., . . . Tassé, M. J. (2016). *Supports Intensity Scale—Children’s Version: User’s manual*. Washington, DC: American Association on Intellectual and Developmental Disabilities.
- Thompson, J. R., Wehmeyer, M. L., Hughes, C., Shogren, K. A., Palmer, S. B., & Seo, H. (2014). The Supports Intensity Scale—Children’s version. *Inclusion, 2*, 140–149. <http://dx.doi.org/10.1352/2326-6988-2.2.140>
- Tremblay, A., & Morin, D. (2015). Use of a psychometric instrument as a referral process for the required level of specialization of health and social services, *Journal of Policy and Practice in Intellectual Disabilities, 12*, 3–11. <http://dx.doi.org/10.1111/jppi.12096>
- van Loon, J., Van Hove, G., Schalock, R., & Claes, C. (2015). Personal outcomes scale for adults. Gent, Belgium: Stichting Arduin and University of Gent.
- Verdugo, M. A., 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*, 496–503. <http://dx.doi.org/10.1352/1944-7558-115.6.496>
- Verdugo, M. A., Guillen, V. M., Arias, B., Vicente, E., & Badia, M. (2016). Confirmatory factor analysis of the Supports Intensity Scale for Children. *Research in Developmental Disabilities, 49-50*, 140–152. <http://dx.doi.org/10.1016/j.ridd.2015.11.022>
- Walker, V. L., DeSpain, S. N., Thompson, J. R., & Hughes, C. (2014). Assessment and planning in K-12 schools: A social-ecological approach. *Inclusion, 2*(2), 125–139. <http://dx.doi.org/10.1352/2326-6988-2.2.125>
- Ward, T., & Stewart, C. (2008). Putting human rights into practice with people with an intellectual disability. *Journal of Developmental and Physical Disabilities, 20*(3), 297–311. <http://dx.doi.org/10.1007/s10882-008-9098-4>
- Wehmeyer, M. L., Buntinx, W. H. E., Lachapelle, Y., Luckasson, R. A., Schalock, R. L., Verdugo, M. A., . . . Yeager, M. H. (2008). The intellectual disability construct and its relation to human functioning. *Intellectual and Developmental Disabilities, 46*(4), 311–318. [http://dx.doi.org/10.1352/1934-9556\(2008\)46\[311:TIDCAI\]2.0.CO;2](http://dx.doi.org/10.1352/1934-9556(2008)46[311:TIDCAI]2.0.CO;2)
- Wehmeyer, M., Chapman, T. E., Little, T. D., Thompson, J. R., Schalock, R., & Tassé, M. J. (2009). Efficacy of the Supports Intensity Scale (SIS) to predict extraordinary support needs. *American Journal on Intellectual and Developmental Disabilities, 114*(1), 3–14. <http://dx.doi.org/10.1352/2009.114:3-14>
- 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*(5), 933–941. <http://dx.doi.org/10.1016/j.ridd.2009.01.007>

Appendix

SIS—A and SIS—C References (2002–2017)

- Arnelsson, G., & Sigurdsson, T. (2014). The validity of the Supports Intensity Scale for adults with psychiatric disabilities. *Research in Developmental Disabilities, 35*, 3665–3671. <http://doi.org/10.1016/j.ridd.2014.09.006>
- Arnelsson, G., & Sigurdsson, T. (2016). The validity of the Supports Intensity Scale for adults with motor disability. *American Journal of Intellectual and Developmental Disabilities, 121*(2), 139–150. <http://dx.doi.org/10.1352/1944-7558-121.2.139>
- Bossaert, G., Kuppens, S., Buntinx, W., Molleman, C., Van den Abeele, A., & Maes, B. (2009). Usefulness of the Supports Intensity Scale (SIS) for persons with other than intellectual disabilities. *Research in Developmental Disabilities, 30*(6), 1306–1316. <http://dx.doi.org/10.1016/j.ridd.2009.05.007>
- Brown, H. K., Ouellette-Kunz, H., Bielska, I., & Elliott, D. (2009). Choosing a measure of support need: Implications for research and policy. *Journal of Intellectual Disability Research, 53*(11), 949–954. <http://dx.doi.org/10.1111/j.13652788.2009.01216.x>
- Chou, Y. C., Lee, Y. C., Chang, S. C., Yu, A. P. (2013). Evaluating the Supports Intensity Scale as a potential assessment instrument for resource allocation for persons with intellectual disability. *Research in Developmental Disabilities: A Multidisciplinary Journal, 34*(6), 2056–2063. <http://dx.doi.org/10.1016/j.ridd.2013.03.013>
- Claes, C., Van Hove, G., Van Loon, J., Vandavelde, S., & Schalock, R. L. (2009). Evaluating the inter-respondent (consumer vs. staff) reliability and construct validity (SIS vs. Vineland) of the supports intensity scale on a Dutch sample. *Journal of Intellectual Disability Research, 53*, 329–338. <http://dx.doi.org/10.1111/j.1365-2788.2008.01149.x>
- Claes, C., Van Hove, G., Vandavelde, S., van Loon, J., Schalock, R. (2012). The influence of support strategies, environmental factors, and client characteristics on quality of life-related personal outcomes. *Research in Developmental Disabilities, 33*, 96–103. <http://dx.doi.org/10.1016/j.ridd.2011.08.024>
- Cruz, M., Perez, M. D., Jenaro, C., Flores, N., & Vega, V. (2013). Identification of the support needs of individuals with severe mental illness using the supports intensity scale. *Revista latino-americana de enfermagem, 21*(5), 1137–1143. <http://dx.doi.org/10.1590/S0104-11692013000500017>
- Cruz, M., Jenaro, C., Perez, M. D., & Robaina, N. F. (2010). Applicability of the Spanish Version of the Supports Intensity Scale (SIS) in the Mexican population with severe mental illness. *Revista latino-americana de enfermagem, 18*, 975–982.

- Giné, C., Adam, A. L., Font, J., Salvador-Bertran, F., Baques, N., Oliveira, C., . . . Thompson, J. R. (2017). Examining measurement invariance and differences in age cohorts on the Supports Intensity Scale—Children’s Version—Catalan Translation. *American Journal on Intellectual and Developmental Disabilities, 122*, 511–524. <http://dx.doi.org/10.1352/19447558122.6.511>
- Giné, C., Font, J., Guardia-Olmos, J., Balcells-Balcells, A., Valls, J., & Carbo-Carrete, M. (2014). Using the SIS to better align the funding of residential services to assessed support needs. *Research in Developmental Disabilities, 35*(5), 1144–1151. <http://dx.doi.org/10.1016/j.ridd.2014.01.028>
- Guillen, V. M., Adam, A. L., Verdugo, M. A., & Giné, C. (2017). Comparisons between the Spanish and Catalan versions of the Supports Intensity Scale for Children (SIS—C). *Psicothema, 29*, 126–132. <http://dx.doi.org/10.7334/psicothema2016.200>
- Guillen, V. M., Verdugo, M. A., Arias, B., & Vicente, E. (2015). Development of a support needs assessment scale for children and adolescents with intellectual disabilities. *Anales de Psicología, 31*, 137–144.
- Guscia, R., Harries, J., Kirby, N., Nettelbeck, T., & Taplin, J. (2006). Construct and criterion validities of the service need assessment profile (SNAP): A measure of support for people with disabilities. *Journal of Intellectual and Developmental Disability, 31*, 148–155. <http://dx.doi.org/10.1080/13668250600876459>
- Harries, J., Guscia, R., Kirby, N., Nettelbeck, T., & Taplin, J. (2005). Support needs and adaptive behaviors. *American Journal on Mental Retardation, 110*(5), 393–404. [http://dx.doi.org/10.1352/0895-8017\(2005\)110%5B393:SNAAB%5D2.0.CO;2](http://dx.doi.org/10.1352/0895-8017(2005)110%5B393:SNAAB%5D2.0.CO;2)
- Jenaro, C., Cruz, M., Perez, M. D, Flores, N. E., & Vega, V. (2011). Utilization of the Supports Intensity Scale with psychiatric populations: Psychometric properties and utility for service delivery planning. *Archives of Psychiatric Nursing, 25*(5), e9–e17. <http://dx.doi.org/10.1016/j.apnu.2011.05.002>
- Kuppens, S., Bossaert, G., Buntinx, W., Molleman, C., Van den Abbeele, A., & Maes, B. (2010). Factorial validity of the Supports Intensity Scale (SIS). *American Journal on Intellectual and Developmental Disabilities, 115*(4), 327–339. <http://dx.doi.org/10.1352/1944-7558-115.4.327>
- Lamoureux-Hebert, M., & Morin, D. (2009). Translation and cultural adaptation of the Supports Intensity Scale in French. *American Journal on Intellectual and Developmental Disabilities, 114*(1), 61–66. <http://dx.doi.org/10.1352/2009.114:61-66>
- Lamoureux-Hebert, M., Morin, D., & Crocker, A. (2010). Support needs of individuals with mild and moderate intellectual disabilities and challenging behaviors. *Journal of Mental Health Research in Intellectual Disabilities, 3*, 67–84. <http://dx.doi.org/10.1080/19315861003650558>
- Lombardi, M., Croce, L., Claes, C., Vandeveld, S., & Schalock, R. L. (2016). Factors predicting quality of life for people with intellectual disability: Results from the ANFFAS study in Italy [Special issue]. *Journal of Intellectual and Developmental Disabilities, 338–347*. <http://dx.doi.org/10.3109/13668250.2016.1223281>

- Morin, D., & Cobigo, V. (2009). Reliability of the Supports Intensity Scale (French version). *Intellectual and Developmental Disabilities, 47*(1), 24–30. <http://dx.doi.org/10.1352/2009.47:24-30>
- Seo, H., Shogren, K. A., Little, T. D., Thompson, J. R., & Wehmeyer, M. L. (2016). Construct validation of the Supports Intensity Scale—Children and Adult versions: An application of a pseudo multitrait-multimethod approach. *American Journal on Intellectual and Developmental Disabilities, 121*, 550–563. <http://dx.doi.org/10.1352/1944-7558-121.6.550>
- Seo, H., Wehmeyer, M. L., Shogren, K. A., Hughes, C., Thompson, J. R., Little, T. D., & Palmer, S. B. (2017). Examining underlying relationships between the Supports Intensity Scale—Adult version and the Supports Intensity Scale—Children’s version. *Assessment for Effective Intervention, 42*, 237–247. <http://dx.doi.org/10.1177/1534508417705084>
- Seo, H., Shogren, K. A., Wehmeyer, M. L., Hughes, C., Thompson, J. R., Little, T. D., & Palmer, S. P. (2016). Exploring shared measurement properties and score comparability between two version of the Supports Intensity Scale. *Career Development and Transition for Exceptional Individuals, 39*, 216–226. <http://dx.doi.org/10.1177/2165143415583499>
- Seo, H., Shogren, K. A., Wehmeyer, M. L., Little, T. D., & Palmer, S. B. (2017). The impact of medical and behavioral support needs on the supports needed by adolescents with intellectual disability to fully participate in community life. *American Journal on Intellectual and Developmental Disabilities, 122*, 173–191. <http://dx.doi.org/10.1352/19447558122.2.173>
- Shogren, K. A., Wehmeyer, M. L., Seo, H., Thompson, J. R., Schalock, R. L., Hughes, C., ... Palmer, S. B. (2017). Examining the reliability and validity of the Supports Intensity Scale – Children’s version in children with autism and intellectual disability. *Focus on Autism and Other Developmental Disabilities, 32*, 293–304. <http://dx.doi.org/10.1177/1088357615625060>
- Shogren, K. A., Seo, H., Wehmeyer, M. L., Palmer, S. B., Thompson, J. R., Hughes, C., & Little, T. D. (2015). Support needs of children with intellectual and developmental disabilities: Age-related implications for assessment. *Psychology in the Schools, 52*, 874–891. <http://dx.doi.org/10.1002/pits.21863>
- Shogren, K. A., Seo, H. S., Wehmeyer, M. L., Thompson, J. R., & Little, T. D. (2016). Impact of the protection and advocacy subscale on the factorial validity of the Supports Intensity Scale—Adult Version. *American Journal on Intellectual and Developmental Disabilities, 121*, 48–64. <http://dx.doi.org/10.1352/1944-7558-121.1.48>
- Shogren, K. A., Shaw, L. A., Wehmeyer, M. L., Thompson, J. R., Lang, K. M., Tassé, M. J., & Schalock, R. L. (2017). The support needs of children with intellectual disability and autism: Implications for supports planning and subgroup classification. *Journal of Autism and Developmental Disorders, 47*, 865–877. <http://dx.doi.org/10.1007/s10803-016-2995-y>

- Shogren, K. A., Thompson, J. R., Wehmeyer, M., Chapman, T., Tassé, M. J., & McLaughlin, C. A. (2014). Reliability and validity of the Supplemental Protection and Advocacy Scale of the Supports Intensity Scale. *Inclusion, 2*, 100–109. <http://dx.doi.org/10.1352/2326-6988-2.2.125>
- Simões, C., Santos, S., Biscaia, R., & Thompson, J. R. (2016). Understanding the relationship between quality of life, adaptive behavior, and support needs. *Journal of Developmental and Physical Disabilities, 28*, 849–870. <http://dx.doi.org/10.1007/s10882-016-9514-0>
- Smit, W., Sabbe, B., & Prinzie, P. (2011). Reliability and validity of the Supports Intensity Scale (SIS) measured in adults with physical disabilities. *Journal of Developmental & Physical Disabilities, 23*, 277–287. <http://dx.doi.org/10.1007/s10882-011-9227-3>
- Soltani, S., Kamali, M., Chabok, A., & Ashayeri, H. (2013). Evaluating the validity and reliability of Persian Version of Supports Intensity Scale in adults with intellectual disability. *Journal of Kermanshah University of Medical Sciences, 17*(9), 555–562.
- Tassé, M. J., & Wehmeyer, M. L. (2010). Intensity of support needs in relation to co-occurring psychiatric disorders. *Exceptionality, 18*, 182–192. <http://dx.doi.org/10.1080/09362835.2010.513922>
- Thompson, J. R., Hughes, C., Schalock, R. L., Silverman, W., Tassé, M. J., Bryant, B., . . . Campbell, E. M. (2002). Integrating supports in assessment and planning. *Mental Retardation, 40*(5), 390–405. [http://dx.doi.org/10.1352/0047-6765\(2002\)040%3C0390:ISIAAP%3E2.0.CO;2](http://dx.doi.org/10.1352/0047-6765(2002)040%3C0390:ISIAAP%3E2.0.CO;2)
- Thompson, J. R., Shogren, K. A., Seo, H., Wehmeyer, M. L., & Lang, K. M. (2016). Creating a SIS—A annual review protocol to determine the need for reassessment. *Intellectual and Developmental Disabilities, 54*, 217–228. <http://dx.doi.org/10.1352/1934-9556-54.3.217>
- 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.
- Thompson, J. R., Wehmeyer, M. L., Hughes, C., Shogren, K. A., Palmer, S. B., & Seo, H. (2014). The Supports Intensity Scale—Children’s version. *Inclusion, 2*, 140–149. <http://dx.doi.org/10.1352/2326-6988-2.2.140>
- Tremblay, A., & Morin, D. (2015). Use of a psychometric instrument as a referral process for the required level of specialization of health and social services. *Journal of Policy and Practice in Intellectual Disabilities, 12*, 3–11. <http://dx.doi.org/10.1111/jppi.12096>
- Verdugo, M. A., Arias, B., Guillen, V. M., Seo, H., Shogren, K. A., Shaw, L. A., & Thompson, J. R. (2016). Examining age-related differences in support needs on the Supports Intensity Scale—Children’s Version—Spanish translation. *International Journal of Clinical and Health Psychology, 16*, 306–314. <http://dx.doi.org/10.1016/j.ijchp.2016.06.002>

- Verdugo, M. A., 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*, 496–503. <http://dx.doi.org/10.1352/1944-7558-115.6.496>
- Verdugo, M. A., Guillen, V. M., Arias, B., Vicente, E., & Badia, M. (2016). Confirmatory factor analysis of the Supports Intensity Scale for Children. *Research in Developmental Disabilities, 49-50*, 140–152. <http://dx.doi.org/10.1016/j.ridd.2015.11.022>
- Wehmeyer, M., Chapman, T. E., Little, T. D., Thompson, J. R., Schalock, R., & Tassé, M. J. (2009). Efficacy of the Supports Intensity Scale (SIS) to predict extraordinary support needs. *American Journal on Intellectual and Developmental Disabilities, 114*(1), 3–14. <http://dx.doi.org/10.1352/2009.114:3-14>
- Weiss, J. A., Lunsy, 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*(5), 933–941. <http://dx.doi.org/10.1016/j.ridd.2009.01.007>