RESEARCH BRIEF
October 2020

Comparing *SIS–A* Assessments Administered Using Face-to-Face and Virtual Interviewing Formats

James R. Thompson & Sarah R. Carlson
University of Kansas

Leslie Shaw
Cornell University

Abstract

The COVID-19 pandemic prompted *Supports Intensity Scale—Adult Version (SIS–A)* (Thompson et al., 2015)® interviewers from multiple U.S. states to conduct assessments using virtual interviews instead of face-to-face (f2f) interviews. To investigate the impact that administration mode had on *SIS—A* scores, *SIS—A* results from assessments conducted virtually during April/May of 2020 were compared to *SIS—A* results from assessments conducted f2f in the same jurisdictions during April/May of 2019. Although differences in the scores from the two time periods were detectable, with few exceptions the magnitude of the differences was small and not meaningful. Therefore, jurisdictions can have confidence in the results of *SIS—A* assessments that have been conducted virtually. However, future researchers should continue to investigate the impact assessment mode has on both the process and outcomes of *SIS—A* assessments.

aaidd
American Association on Intellectual and Developmental Disabilities

Background

The *Supports Intensity Scale—Adult Version (SIS—A)* was first published by the American Association on Intellectual and Developmental Disabilities (AAIDD) in 2004. It was designed to provide a psychometrically valid means to assess and measure the intensity of support needed by people with intellectual and developmental disabilities (IDD). When a children's version of the scale (i.e., the *SIS–C*®) was created for field testing, the *SIS—A* was refreshed and renamed in 2015 (it was simply known as the Supports Intensity Scale prior to 2015). The *SIS—A* has been translated into 14 languages and used in multiple countries for a variety of purposes such as determining program eligibility, planning supports and services, and distributing funds (AAIDD; 2020a; Schalock et al., 2018). Its widest use has been in North America with 26 U.S. states or Canadian provinces using it for numerous purposes since it was first published (AAIDD, 2020e). In some jurisdictions it has been used only with a small population for the purpose of planning personalized supports. Other jurisdictions have used it more extensively, including for the relatively "high stakes" purpose of informing the development of supports budgets. This includes determining how Medicaid funds in Home and Community Based Services [HCBS] programs for long-term services and supports [LTSS] should be distributed (see Agosta et al., 2016; Virginia Department of Behavioral Health & Developmental Services, 2014).

Of the standardized support needs assessment scales that have been published during the past 20 years, the *SIS—A* is by far the most well-known and widely used (Thompson & DeSpain, 2016). Extensive research findings have supported its psychometric soundness. In a comprehensive review of peer-reviewed published literature regarding the *SIS—A*, Thompson et al. (2018) found strong support for indicators of reliability (i.e., internal consistency, split-half, test-retest, interrater) and validity (i.e., content, criterion-related, construct). Moreover, research on translated versions of the *SIS—A* has shown the scale's psychometric properties remain robust across different languages and cultures (e.g., Arnkelsson & Sigurdsson, 2014; Chou et al., 2013; Lombardi et al., 2016; Simeos et al., 2016).

Like any other assessment scale, information gleaned from a *SIS—A* assessment is only as valid as the extent to which the assessment was completed correctly. Because the *SIS—A* is completed through a structured interview with respondents who know the person being assessed, *SIS—A* assessors need to know when to ask follow-up questions and probe respondents in order to arrive at accurate ratings (Thompson et al., 2015). When *SIS—A* assessments are used by jurisdictions to inform the development of supports budgets and therefore allocate public funds, the assessment is considered to be high stakes. As the publisher of the *SIS—A*, the AAIDD has insisted that when the *SIS—A* is used for high stakes purposes, it must be conducted by assessors who have completed interviewer training that meets the AAIDD's national standards on how to administer it. The AAIDD (2020c) has developed a comprehensive training and qualification program that culminates in assessors demonstrating proficiency in its administration.

Because *SIS—A* assessors have been trained to conduct face-to-face (f2f) interviews and because data for the standardization sample were collected through f2f interviews, jurisdictions using AAIDD certified, qualified, and recognized assessors have conducted *SIS—A* assessments in person. With the advent of the COVID-19

pandemic, conducting f2f assessments became dangerous to the health and safety of assessors and respondents. Jurisdictions were faced with several choices, none of which were ideal. For instance, they could continue with f2f interviews and implement as many precautions (e.g., social distancing during interviews, masks worn by all parties) as feasible. However, precautions are not fool-proof and conducting f2f *SIS—A* interviews introduces an element of risk no matter what safety protocols are used. Alternatively, they could discontinue *SIS—A* assessments until the pandemic passed. However, that would result in the provision of LTSS being delayed for people new to a HCBS program as well as potentially inequitable allocation of HCBS funding for people enrolled in the program (because other people who had been enrolled had periodically been reassessed and their funding levels adjusted accordingly). A third option was to conduct the *SIS—A* assessments through virtual interviews using video conferencing platforms (e.g., ZOOM, MS Teams, WebEx, or Skype) or telephone calls. However, as alluded to earlier, such assessments would mean that information was collected differently from others in their HCBS program in addition to being different from the assessments used to establish the norms for the *SIS—A*. The final option was a combination of all three. Namely, to postpone some assessments, conduct some assessments f2f, and conduct other assessments virtually.

The purpose of this study was to investigate the impact that administration mode (i.e., f2f versus virtual) had on *SIS—A* scores. Nine U.S. states verified that their *SIS—A* **results from April and/or May of 2020 were from assessments conducted virtually**, and their *SIS—A* **results from April and/or May of 2019 were from assessments conducted f2f**. It is logical to assume that the sample of people from a jurisdiction, who were enrolled in the same HCBS program, would be very similar from one year to another. Thus, it is highly likely that the aggregate support needs of these two groups would be very similar. Therefore, comparing *SIS—A* results from assessments conducted under one condition in 2019 (i.e., f2f interviews) with those conducted under a different condition in 2020 (i.e., virtual interviews) can provide insight into differences in *SIS—A* results that might be attributable to the different assessment administration modes.

> "The purpose of this study was to investigate the impact that administration mode (i.e., f2f versus virtual) had on *SIS—A* scores."

Method

Research Question

The research question guiding this investigation was straightforward: "Are there statistically significant and/or practical differences between *SIS—A* scores when assessments are conducted face-to-face versus virtually?"

Participants

The participants were people who had been determined eligible for services through a state HCBS 1915(c) Waiver program, which is the major Medicaid program that pays for LTSS for people with IDD. As part of the process to develop supports budgets (i.e., distribute funding) through the HCBS program, all participants were assessed with the *SIS—A*. Their assessment results were entered into the *SISOnline* (AAIDD, 2020d), a data repository maintained by the AAIDD.

The Institutional Review Board at the University of Kansas granted approval to the researchers to access deidentified data from the *SISOnline* and AAIDD provided

the deidentified data set for this analysis. As Table 1 shows, *SIS—A* results on 3,753 people were collected in April–May 2019 which was the f2f condition, and *SIS—A* results on 2,862 people were collected in April–May 2020 which was the virtual condition.

The participants came from nine geographically diverse U.S. states. The diversity of states is important to note because HCBS programs are not uniform across states. In fact, there is significant variability in the populations of people with IDD from different states who are determined to be eligible for and become enrolled in HCBS programs, as well as the amount of funding that is made available for distribution (Braddock et al., 2019). Thus, having data from nine states provides data from a greater cross-section of people with IDD than if data from only one or a few states were included. Two states, collected virtual *SIS—A* data only in May 2020. Therefore, data from April of 2019 and 2020 from these two states were excluded from the analysis. The remaining states collected virtual data in both April and May 2020.

Section 1:
Exceptional Medical and Behavioral Support Needs
(32 items)

provides a means to measure a person's additional support needs due to exceptional medical conditions and/or behavioral concerns

Section 2:
Support Needs Index
(49 items)

describes life activities and is rated against 3 support dimensions (frequency, type, and time)

Section 3:
Supplemental Protection and Advocacy Scale
(8 items)

is focused on support needed for protection and advocacy activities but does not generate standard scores

Materials

All participants were assessed with the *SIS—A*. The *SIS—A* is made up of three sections. Section 1 (32 items) provides a means to measure a person's additional support needs due to exceptional medical conditions and/or behavioral concerns. The standardized portion of the *SIS—A*, Section 2, is comprised of 49 items distributed across six subscales. Each item describes a life activity and is rated against 3 support dimensions (frequency, type, and time). Each dimension is rated on a 5-point scale. Upon completion of a *SIS—A* assessment, two norm-referenced indices of support needs are generated: (1) the Support Needs Index (a composite score that reflects a person's overall intensity of support needs) and (2) the Support Needs Profile (a graphic plot that shows a pattern of a person's intensity of support needs across six support need domains/subscales). *SIS—A* standard scores indicate the relative intensity of people's support needs in relation to a representative sample of adults with IDD. The final section of the *SIS—A*, Section 3 (8 items), is focused on support needed for protection and advocacy activities but does not generate standard scores. Only the standard scores from the *SIS—A* were used in this analysis.

Procedures

Data were collected by *SIS—A* assessors who were trained and qualified by AAIDD. *SIS—A* administration procedures are outlined in detail in the *SIS—A User's Manual* (Thompson et al., 2018) and additional instructions are described in AAIDD's *SIS—A* training materials (AAIDD, 2020c). AAIDD (2020b) has reported that *SIS—A* assessments typically take between 2 and 2.5 hours to complete, however no data

were available regarding the time allocated for completing the assessments in the two conditions. Data from 2019 were collected through f2f interviews and data from 2020 were collected through virtual interviews. Because this investigation was a natural experiment (i.e., people were only assigned to the virtual assessment condition because the COVID-19 pandemic placed people participating in f2f assessments at risk), participants were not assigned to conditions by the researchers. Rather, their placement in either the virtual or f2f assessment group was a function of when their jurisdiction scheduled their *SIS—A* assessment.

Analysis

To answer the research question, *t*-tests and confidence interval testing were applied to compare means between the two groups, f2f versus virtual administration. The alpha level was set to .05 for rejection of the null hypotheses. As recommended by the American Psychological Association when comparing means (APA, 2019), Cohen's *d* was calculated to provide a measure of the standardized mean difference between groups that is independent of sample size. Based on absolute values of *d*, *d* of 0.10 – 0.30 is small, 0.30 – 0.50 is medium, and greater than 0.50 is large.

Although Cohen's *d* provides a quantitative measure of magnitude, determining whether there is a meaningful difference in scores depends on the field of inquiry and the assessment that is used. "Meaningful" is a subjective term, so any standard that is established is going to be subjective. Moreover, the boundary between "meaningful and non-meaningful" is going to be arbitrary. Nevertheless, because IDD is an applied field, determining what magnitude of difference is meaningful is every bit as salient to drawing conclusions regarding the potential impact that virtual administration has on *SIS—A* scores as are indicators of statistical difference and effect size. To address this need for a more concrete practical difference, the standard error of the mean (SEm) was selected as a second measure.

"To answer the research question, *t*-tests and confidence interval testing were applied to compare means between the two groups, f2f versus virtual administration."

The SEm estimates how repeated measures of a person on the same instrument would be distributed around a person's "true" score (the true score is always unknown because it is impossible to construct a measurement scale that provides a faultless true score; Salkind, 2010). According to Thompson et al. (2015) the SEm of the *SIS—A* is 1.5, meaning that a person's true score lies somewhere between plus or minus 1.5 of their observed SNI score. Thus, because a person's true score is highly likely to fall within a 3-point window (either 1.5 above or 1.5 below), it is reasonable to suggest that a meaningful difference between mean SNI scores would require that the scores be separated by at least 3 points. As for the subscales, Thompson et al. (2015) reports a SEm range of 0.52 to 0.72, depending on the subscale. Going with the lowest subscale SEm, it is reasonable to suggest that a meaningful difference between mean subscale scores would be at 1.04 points or more.

Results

As Table 1 shows, when data from all nine states were merged and the composite score (i.e. the Support Needs Index or SNI) was used as the dependent variable, statistical differences were found ($t_{6401.9} = -5.13$, $p < .001$) when scores generated from f2f assessment (April-May, 2019) were compared with scores generated through

Table 1

	Face-to-Face			*Virtual*			*Compare*	
	n	*M*	*SD*	*n*	*M*	*SD*	*ΔM*	*d*
All 9 states	3753	100.99	11.71	2862	102.41	10.70	1.42*	0.12
State 1	390	101.86	9.41	231	103.16	8.25	1.30*	0.14
State 2	248	91.58	16.47	119	93.13	14.75	1.55	0.10
State 3	325	99.65	12.75	419	104.36	11.37	4.71*	0.39
State 4	861	103.30	10.55	317	104.98	8.73	1.68*	0.17
State 5	133	97.67	10.66	111	100.19	10.90	2.52*	0.23
State 6	144	90.57	14.93	153	93.63	14.27	3.06*	0.21
State 7	127	104.69	12.08	9	93.78	14.73	–10.91*	–0.89
State 8	597	101.29	10.18	373	103.30	7.87	2.01*	0.21
State 9	928	102.87	9.75	1130	102.97	9.94	0.10	0.01

Note. Δ*M* is equal to virtual mean minus face-to-face mean. * indicates that the means are different based on a 95% confidence interval test that assumes equal variance. *Cohen's d* is represented by *d*. Based on absolute values of *d, d* of 0.10 – 0.30 is small, 0.30 – 0.50 is medium, and greater than 0.50 is large.

virtual assessment (April-May, 2020). Although the mean scores collected during the two administrations were statistically different at the 95% confidence level, the numerical difference between the mean scores was small (i.e., M_{diff} = 1.42).

When SNI scores were examined by state, differences between mean scores from f2f and virtual assessments ranged from –10.91 for State #7 to 4.71 for State #3. For the other seven states, the mean score differences were much smaller in absolute value ranging from 0.10 to 2.52. In seven of the nine states, the mean score differences were statistically different at the 95% confidence interval, with State #9 and State #2 being the two exceptions.

State #7 had a negative mean difference, which indicated that a higher average SNI score (i.e., more intense support needs) was present during f2f administration. The other eight states and the merged data (i.e., all states together) showed positive mean differences, which indicated higher average SNI scores (i.e., more intense support needs) during virtual administration.

Table 2 shows findings when the dependent variables were the subscale scores. The scores for all six subscales were statistically

Table 2
Subscale standard score descriptive statistics and mean comparisons

	Face-to-Face			*Virtual*			*Compare*	
Activities	*n*	*M*	*SD*	*n*	*M*	*SD*	*ΔM*	*d*
Home Living	3753	9.77	2.62	2862	10.14	2.52	0.37*	0.14
Community Living	3753	9.13	1.64	2862	9.27	1.49	0.14*	0.09
Lifelong Learning	3753	11.41	1.72	2862	11.55	1.62	0.14*	0.08
Employment	3753	10.16	1.59	2862	10.32	1.38	0.16*	0.11
Health and Safety	3753	10.41	1.97	2862	10.66	1.85	0.25*	0.13
Social	3753	10.02	1.89	2862	10.21	1.68	0.19*	0.11

Note. Δ*M* is equal to virtual mean minus face-to-face mean. * indicates that the means are different based on a 95% confidence interval test that assumes equal variance. *Cohen's d* is represented by *d*. Based on absolute values, *d* of 0.10 – 0.30 is small, 0.30 – 0.50 is medium, and greater than 0.50 is large.

different in the two administration conditions. However, the numerical differences between the subscale mean scores were all quite small, ranging from 0.14 to 0.37.

Table 1 shows the effect size was small for the merged data set and all of the states except for State #3 (moderate effect size) and State #7 (large effect size). In State #9, it is noteworthy that the Cohen's *d* effect size was near 0.00, meaning the scores collected during the two conditions were essentially identical. Table 2 shows the effect sizes for the subscales and SNI scores were all very small, with the largest effect size (d = 0.14) for Home-Life Activities and the smallest for Lifelong Learning Activities (d = 0.08).

Using the 3-point standard to guide the interpretation of meaningful differences in SNI means, Table 1 shows the merged mean score difference (1.42) was well within this threshold. Only two of the nine states had mean score differences greater than 3 points. Table 2 shows that none of the subscales had difference scores beyond 0.37; therefore, they all fell well below the 1.04 threshold that was established to denote a meaningful difference.

Discussion

Statistical and Meaningful Differences

The findings from this study strongly suggest that jurisdictions should have confidence in the results of *SIS—A* assessments that were conducted virtually in 2020. Although multiple statistical differences were evident, there were very few statistical differences that were of practical importance (i.e., had an effect size greater than "small" or indicated meaningful differences). The concept of statistical significance can be misused when evidence of statistical differences (i.e., differences not due to chance) is assumed to have practical implications. In this case, statistical differences were primarily due to the large sample size. This is especially important to keep in mind when reviewing findings from this investigation.

The statistical differences identified in Tables 1 and 2 indicate the probability of the differences occurring because of chance. It is highly unlikely (95 of 100 times) that the differences in means that are identified on the two tables as being "statistically different" are the result of a chance occurrence. Rather, these differences are most likely due to some factor of interest. There can, however, be many factors that influence differences in SNI mean scores. How many of the differences were due to assessment administration mode is unknown.

"The findings from this study strongly suggest that jurisdictions should have confidence in the results of *SIS—A* assessments that were conducted virtually in 2020."

The differences in mean scores could have been influenced by the characteristics of the two samples. Although we stated previously that it was logical to assume that the people assessed virtually in 2020 were very similar to the people assessed f2f in 2019, "very similar" does not mean "exactly the same." There were surely some differences between the two samples and these differences may be responsible for the slight (but statistically significant) differences in the mean scores. The fact that so many more people were assessed in 2019 suggests that it may have been more difficult to arrange virtual interviews in 2020 than arrange f2f interviews in 2019. Thus, the mean score differences in the two time periods could be an artifact

associated with absence of people in the 2020 sample for whom virtual interviews could not be arranged.

Other conceivable influences were the different assessors who collected the *SIS—A* data. Although many assessments were completed by the same *SIS—A* assessors in both years, there were undoubtedly some assessors who completed assessments in 2019 but not in 2020, and vice versa. Moreover, the proportion of assessments completed in a state by different assessors surely changed from 2019 to 2020. Having AAIDD trained and qualified assessors is intended to prevent wide variability in *SIS—A* results, but it does not preclude any and all variability.

"The critical point to be made is that statistical differences between mean scores in 2019 and 2020 cannot solely be attributed to differences in assessment mode."

The critical point to be made is that statistical differences between mean scores in 2019 and 2020 cannot solely be attributed to differences in assessment mode. The only knowledge claim that can be made in this investigation is that there were several statistical differences identified in the mean scores from the different time frames. However, the logic underlying the investigation is that changes in the sample and in the assessors, as well as any other variables that one could imagine (e.g., people throughout the population were feeling more stressed in 2020 due to COVID-10, which could impact rapport during the assessments), would only result in small differences in the mean scores from the two time frames. Administration mode, however, could potentially have a much greater influence on mean scores because it changes the way *SIS—A* information is collected. A different mode of information input could result in different output.

Therefore, evidence of statistical significance is the least interesting finding of this investigation and indicators of the magnitude of change (based on Cohens *d* as well as the subjective criteria that was established) are the most interesting . If the differences in *SIS—A* scores during 2019 and 2020 were of a significant magnitude, then the use of different administration modes in those time periods would be concerning even though their relative influence could not immediately be parceled out. However, the findings from this investigation showed the opposite. The magnitude of differences was quite small overall, which suggests that neither the assessment mode nor other potential factors had much of an influence on how the *SIS—A* was operating.

Interpreting Scores from the Outliers

There were two states which had both meaningful and statistical differences between their 2019 and 2020 mean scores. The results from State #7 must be discounted immediately because the state only conducted 9 virtual interviews in 2020 compared to 127 f2f interviews in 2019. The sparse numbers in 2020 make any conclusion highly spurious; for instance, the results could be completely different if just five more people would have been assessed in 2020. Although this state could have been removed from the analysis due to its low numbers, it met the inclusion criteria when it was confirmed that their assessors had shifted to virtual assessment in 2020. There was no justification for dropping it simply because the numbers were small. State #7's data still contributed (albeit in a meager way in 2020) to the richness of the merged data set.

State #3 is a different case. In this state the number of participants in both years was

large (325 in 2019 and 419 in 2020), and the mean scores associated with the two assessment periods differed by 4.71 points. This increase in *SIS—A* SNI scores from 2019 to 2020 not only exceeds the 3.0 SEm-based threshold that was set as a standard to denote a meaningful difference, but does so by an unambiguous margin. Moreover, Cohen's *d* indicates the effect size of 0.39 is in the moderate range. Although the difference in mean SNI scores between 2019 and 2020 in State #3 is not so large as to cast doubts on the entire *SIS—A* assessment process, further investigation into potential reasons why *SIS—A* scores are higher in 2020 than 2019 is warranted. If there is a suspicion that the mode of assessment is contributing to the discrepancy, State #3 should consider implementing further quality control measures of their assessment practices, as well as training in how to most effectively conduct virtual assessments.

Limitations

Several limitations of this analysis have been alluded to in this paper, and statistical best practices were applied to the extent-possible to counter those limitations. The first limitation was sample size exceeding 1,000, leading to almost any *t*-test returning statistically significant results due to small standard errors. To counter this limitation, Cohen's *d* was calculated from group means and standard deviations providing values that could be interpreted as practical differences. SEm values were also set for the SNI and subscales to determine which state's scores had changed to a meaningful degree. Another limitation was the small sample size of virtual assessments from State #7. The results from that state may have skewed the *t*-test results, but those observations met inclusion criteria for the sample so they were retained. Lastly, other characteristics of the sample were not included as covariates or control variables. This analysis was based on the assumption that both samples were equivalent on all relevant characteristics, and there surely were some noteworthy differences between the two samples.

A final limitation of this research study was the reliance on the SNI and standardized subscale scores. To obtain the SNI and standardized subscale scores, raw scores are tabulated. The SNI and subscale scores are obtained by finding the values corresponding to the raw scores in the norming tables. In that process, information about individual questions is lost, information that could provide insight into how the measure performs under the two conditions, f2f versus virtual. Future analyses are needed which allows the evaluation of individual items in a confirmatory factor analysis (Brown, 2006) framework.

Future Research

More research is needed to best understand how the shift from f2f to virtual assessment might impact the administration and scoring of the *SIS—A*. As mentioned, to extend this research line it is critical to evaluate data in a confirmatory factor analysis (Brown, 2006) framework to allow both measurement properties and latent properties of the *SIS—A* to be tested. It is equally important to include covariates in future studies. Adding covariates to the analysis will control for group differences between those assessed f2f versus virtually. Such analysis is essential to determining ways in which mean score differences could be attributed to sample characteristics

As important as it is to determine the correct support needs scores when administering the *SIS—A*, this is not the assessment's only use. For example, the f2f interview associated with *SIS—A* administration provides a time for people with IDD, their families, and professionals to generate information to inform decision-making and person-centered planning. Additionally, the f2f

interview associated with *SIS—A* administration allows for rapport to be established between the *SIS—A* assessor and respondents, which in turn influences respondents' confidence in the assessment results, and by extension the system of LTSS as a whole. Considering this, future research needs to assess the impact of *SIS—A* administration format changes on the assessment's secondary uses. To best address these topics, researchers should focus on understanding the perspectives and experiences of *SIS—A* assessors who have experience with both f2f and virtual administration formats. Such research can provide a nuanced understanding of how the administration format might impact the stability of *SIS—A* scores; inform changes to *SIS—A* assessor training guidelines and materials; support state-level officials to make decisions regarding the use of the virtual *SIS—A* administration format; and inform future research on this topic. Also, gathering qualitative data on the perspectives of people with disabilities and their family members regarding the assessment process is essential.

Conclusion

Findings from the current study provide assurances to jurisdictions wishing to continue to conduct *SIS—A* assessment virtually for as long as the COVID-19 pandemic presents safety concerns. Although there is no reason to not continue with virtual assessments, it is important for jurisdictions to monitor their *SIS—A* assessment results and take appropriate action if results from virtual assessments differ meaningfully from results obtained from f2f assessments. Additional research is need to provide a more comprehensive understanding of the impact that administration mode has on *SIS—A* results.

References

Agosta, J., Petner-Arrey, J., Aiken, F., Kardell, Y., Vazquez, A., Rojas, R., ... Villwock, M. (2015). *Information Brief: Person-centered planning and supports budgets.* Human Research Institute. http://www.hsri.org/publication/information-brief-person-centered-planning-and-supports-budgets

American Association on Intellectual and Developmental Disabilities (AAIDD; 2020a). *International SIS Use*. Retrieved from http://aaidd.org/sis/international#.UsWhcPaSKW4

American Association on Intellectual and Developmental Disabilities (AAIDD; 2020b). *Most frequently asked questions about SIS*. Retrieved from https://www.aaidd.org/sis/product-information/faqs

American Association on Intellectual and Developmental Disabilities (AAIDD; 2020c). *Overview of SIS Training Programs*. Retrieved from https://www.aaidd.org/sis/training/overview

American Association on Intellectual and Developmental Disabilities (AAIDD; 2020d). *SISOnline. https://www.aaidd.org/sis/sisonline*

American Association on Intellectual and Developmental Disabilities (AAIDD; 2020e). *States/Provinces using SIS in North America*. Retrieved from https://www.aaidd.org/sis/sisonline/states-using-sis

American Psychological Association (APA; 2019). Publication manual of the American Psychological Association (7th ed.).https://doi.org/10.1037/0000165-000

Arnkelsson, G., & Sigurdsson, T. (2014). The validity of the Supports Intensity Scale for adults with psychiatric disabilities. *Research in Developmental Disabilities, 35*, 3665–3671. http://dx.doi.org/10.1016/j.ridd.2014.09.006

Brown, T. A. (2006). *Confirmatory factor analysis.* Guilford Press.

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*, 2056–2063.

Lombardi, M., Croce, L., Claes, C., Vandevelde, S., & Schalock, R. L. (2016). Factors predicting quality of life for people with intellectual disability: Results from the ANFFAS study in Italy. *Journal of Intellectual and Developmental Disabilities,* http://dx.doi.org/10.3109/13668250.2016.1223281

Salkind, N. J. (2010).*Encyclopedia of research design.* SAGE Publications. http://dx.doi.org/10.4135/9781412961288.n436

Schalock, R. L, Thompson, J. R., & Tassé, M. J. (2018). *How organizations and systems use Supports Intensity Scales.* American Association on Intellectual and Developmental Disabilities. https://aaidd.org/sis/white-papers#.Wvr6V6kh1o5

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

Thompson, J. R., Bryant, B., Schalock, R. L., Shogren, K. A., Tassé, M.J., Wehmeyer, M. L., Campbell, E. M., Craig, E. M., Hughes, C., & Rotholz, D. A. (2015). *Supports Intensity Scale – Adult Version*: *Users Manual.* American Association on Intellectual and Developmental Disabilities.

Thompson, J. R., & DeSpain, S. N. (2016). Community support needs. In N. N. Singh (Ed.), *Handbook of Evidence-based Practices in Intellectual and Developmental Disabilities (pp.* 137-168). Springer.

Virginia Department of Behavioral Health & Developmental Services. (2014). *Information Brief: Supports Intensity Scale and assessment levels.* http://www.vaaccses.org/vendorimages/vaaccses/BRIEF_SISandAssessmentLevels_Cochran_111314.pdf

Suggested citation: Thompson, J.R., Carlson, S.R. & Shaw, L. (2020, October). *Comparing* SIS—A *assessments administered using face-to-face and virtual interviewing formats.* American Association on Intellectual and Developmental Disabilities.

For questions or more information, please contact us at https://www.aaidd.org/sis/contact-us or at AAIDD, 8403 Colesville Road, Suite 900, Silver Spring, MD 20910.