Meeting the COVID-19 Deadlines: Choosing Assessments to Determine Eligibility

Shelley Kathleen Krach[1] , **Tracy L. Paskiewicz[2]**,
Staci C. Ballard[2], **James E. Howell[1]**,
and Suzanne M. Botana[3]

Abstract
Timely identification of children with disabilities is required by federal special education law (Individuals with Disabilities Education Improvement Act, 20 U.S.C. § 1400, 2004). During COVID-19, school psychologists have been faced with the challenge of completing valid, comprehensive, and diagnostic assessments when traditional methods are not an option. Traditional methods of testing have become nearly impossible due to social distancing requirements; therefore, alternate methods need to be considered. These alternate methods may be unfamiliar to the practitioner and/or lack validation to use with confidence. This study offers a prospective guide to help practitioners make safe and valid test selection and interpretation decisions during a pandemic. Examples of assessments analyzed using this guide are provided for the reader. In addition, a case study is provided as an example.

Keywords
assessment, special education policy, legal and ethical issues, technology, COVID-19

Many school psychologists have been unable to complete comprehensive, face-to-face assessments during the COVID-19 pandemic (U.S. Department of Education, 2020). Without mechanisms in place to determine eligibility in a timely manner, schools are at risk of failing to meet the national child find mandates set forth in IDEIA (2004). During COVID-19, the practice of traditional face-to-face testing placed the health of students, families, and practitioners at risk (NASP, 2020a, 2020b). School psychologists have been stuck between the competing legal requirements of making valid, diagnostic decisions, and the health risks associated with contagion.

During spring 2020, direct psychological testing was largely put on hold while school psychologists sought guidance from various entities (i.e., national and state associations, governmental agencies, and test publishing companies). Direct assessments are those that are given directly to the test taker as opposed to indirect assessments that are given to someone else (e.g.,

[1]Florida State University, FL, USA
[2]University of Massachusetts-Boston, MA, USA
[3]Yarmouth School Department, ME, USA

Corresponding Author:
S. Kathleen Krach, Educational Psychology and Learning Systems, Florida State University, 1114 West Call Street, Tallahassee 32306-4453, FL, USA.
Email: skrach@fsu.edu

teacher or parent) on behalf of the client. Both direct and indirect assessments are typical components of any comprehensive assessment battery (Sattler, 2018).

The Office for Civil Rights (March 16, 2020) stated that all face-to-face testing was to be discontinued until schools reopened. From anecdotal reports, many school psychologists replaced face-to-face testing with record reviews, interviews with key stakeholders, and behavioral rating scales that could easily be adapted to a virtual format. Therefore, diagnoses (i.e., anxiety, depression, and Attention Deficit Hyperactivity Disorder) requiring only indirect or qualitative assessments (e.g., rating scales, review of records, diagnostic interviews, observations, etc.) were possible while schools were closed.

Unlike social, emotional, and behavioral diagnoses, it is nearly impossible to diagnose specific learning disorders/learning disabilities (SLD; Sattler, 2018) using only indirect methods. To diagnose SLD, direct assessment data may be needed to determine the child's overall cognitive ability, cognitive processing strengths and weaknesses, and/or their current academic achievement (Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, APA, 2013; DSM-5, 2013; IDEIA, 2004). Computer-based tele-assessments exist that may help gather some of this data; however, many of them are unfamiliar to practicing school psychologists (Krach & Sattler, 2018). Instead, many publishing companies are adapting their face-to-face assessments to be used through a tele-assessment method (Henke, 2020; PAR, 2020; Walker, Taylor, Wright, & Johnsrud, 2020a, 2020b).

Unfortunately, the use of adapted tele-assessment and/or unfamiliar tele-assessment methods may result in data that are psychometrically problematic. Therefore, it is vital for practitioners to understand the strengths and weaknesses associated with each of these techniques. Being unable to make accurate SLD diagnoses is not a trivial manner. This is because SLD diagnoses make up 33% of all special education cases (U.S. Department of Education, 2019) as well as more than 50% of school psychologists' diagnostic caseloads (Bramlett, Murphy, Johnson, Wallingsford, & Hall, 2002).

As school closures or distance education requirements extend from weeks into months, psychologists are faced with these important questions:

- How should I complete cases that need cognitive/intellectual, processing, and/or academic achievement testing without violating social distancing restrictions?
- What valid methods are available if traditional, face-to-face testing is not safe or possible?
- What issues do I need to consider regarding ethical practice when using alternative or tele-assessment methods?

This study presents one possible decision-making guide for school-based practitioners to use when choosing diagnostic assessments and making special education eligibility decisions.

Development of the Guide

Beginning in March 2020, the COVID-19 crisis forced the closures of schools nationwide. That same month, the United States Department of Education's Office for Civil Rights released a set of guidelines regarding COVID-19 in which they stated, "If an evaluation of a student with a disability requires a face-to-face assessment or observation, the evaluation would need to be delayed until school reopens" (p. 3). Soon after, recommendations for testing during COVID-19 began to emerge from various other entities. Krach, Paskiewicz, and Monk (2020) collected and evaluated these recommendations between May 1 and July 1 of 2020.

They identified three different types of recommending entities: test publishers (e.g., Henke, 2020; PAR, 2020; Walker et al., 2020a, 2020b; etc.), professional organizations (e.g., NASP, 2020a, 2020b; Wright, Mihura, Pade, & McCord, 2020; randomly selected state groups; etc.), and

government agencies (e.g., U.S. Department of Education, 2020; United States Social Security Administration, 2020; randomly selected state agencies; etc.). In total, a sample of 51 different entities was identified in their study.

Krach and colleagues (2020) identified four adapted tele-assessment recommendation trends issued by these entities. *Adapted tele-assessment* was defined as using a telecommunications device to administer tests developed for face-to-face use. These four recommendation trends included: (1) no advice given, (2) recommend to avoid tele-assessment, (3) use tele-assessment with caution, and (4) use tele-assessment without concern. Entities were inconsistent in their recommendations for the use of adapted tele-assessment (Krach et al., 2020), largely depending on whether they were issuers or regulators of the tests. Specifically, test publishers were more likely to recommend the use of tele-assessment (with or without caution), while government agencies were more likely to advice against the use of this method, or to only use it with caution. Professional organizations' recommendations often fell into the "use with caution" category.

Unfortunately, there is no standard for what "use with caution" means regarding adapted tele-assessments. This is in part because the American Psychological Association (APA, 2017) *Ethical Principles of Psychologists and Code of Conduct* was published prior to COVID-19. Standard 9.02 provides some guidance related to how one might use caution in the current crisis. It states the following:

1. Psychologists administer, adapt, score, interpret, or use assessment techniques, interviews, tests, or instruments in a manner and for purposes that are appropriate in light of the research on or evidence of the usefulness and proper application of the techniques.
2. Psychologists use assessment instruments whose validity and reliability have been established for use with members of the population tested. When such validity or reliability has not been established, psychologists describe the strengths and limitations of test results and interpretation.

This means that any definition of "use with caution" must address the following components: (1) the purpose of the assessment, (2) evidence related to the validity of the assessment, (3) evidence related to the reliability of the assessment, and (4) caveats included in any interpretations made from the data collected. The following guidelines presented in this study are designed to help practitioners better navigate the review of each of these components. The goal of this guide is to help practitioners make better and more cautious decisions about current assessment practice. It should be used alongside professional judgment; the guide alone cannot guarantee safe and/or psychometrically sound assessments.

In the following sections of the study, traditional assessments are described as ones that are frequently used in psychoeducational examinations. Sample traditional assessments discussed here were chosen from the two Sattler assessment textbooks (2014, 2018). In contrast, non-traditional assessments are described as tests that were originally developed to be tele-assessments, but they are not included in either of the Sattler texts. Sample nontraditional tele-assessment tests that were selected following a comprehensive internet search. Nontraditional tele-assessments were selected for inclusion in this study if they met the following criteria: (1) originally developed for computer-based use, (2) accompanied by manuals and/or peer-reviewed research, and (3) measured psychoeducational constructs. There were few computer-based assessments that met these requirements. The ones that best met the necessary criteria were included in this study.

The final group of tests discussed in this study includes traditional assessments that were adapted for administration through tele-assessments methods. There is a dearth of conclusive literature on the psychometric properties of these adapted tele-assessments. All available, published, and adapted tele-assessment studies were included in this study.

Please note, the inclusion of a test in the guide/study does not indicate an endorsement by the authors.

Using the Potential Guide

Overview

Figure 1 provides each of the steps in the guide. The *first step* is to identify the type of data needed to make a diagnosis or an eligibility determination (IDEIA, 2004; Miller et al., 2018). For a comprehensive assessment, the type of data includes both qualitative data (e.g., interviews, reviews of records, observations, etc.) and quantitative data (e.g., rating scales, cognitive assessments, achievement tests, etc.).

Once the type of data is selected, the *second step* in this guide is to examine how the data are to be collected. The guide provides three possible color-coded options (i.e., purple/solid, orange/dotted, and blue/dot-dash) to help practitioners evaluate different data collection methods. In the
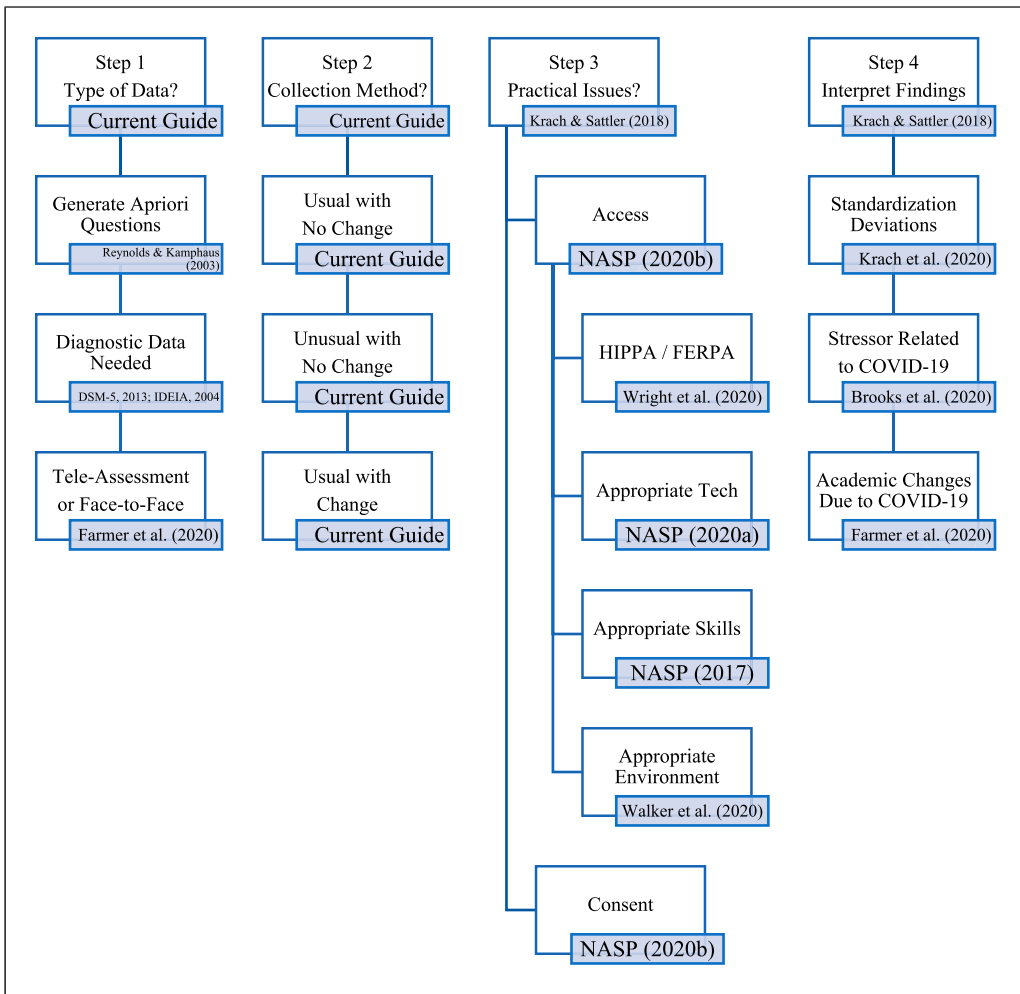

Figure 1. Steps in the COVID-19 assessment decisions.

following sections of this study, the authors will describe the differences between the three options and provide sample analyses of specific tests available.

Finally, the third step requires that the practitioner evaluate legal, ethical, and practical considerations related to the assessment methods chosen. For example, practitioners should consider if examinees have access to digital assessment. Access includes both the availability of the required technology as well as the skills needed to complete computer-based administration or tele-assessment (NASP, 2020b). Practitioners must also maintain test security and confidentiality of client records and responses (Wright et al., 2020). Tele-assessment procedures will also require changes to typical informed consent (NASP, 2020b, Wright et al., 2020).

The *fourth step* addresses interpretation of the assessment findings. In the assessment report, practitioners should explicitly state when modifications to standard procedures were used and provide a rationale for these changes. When interpreting assessment data gathered through nontraditional methods, practitioners should take care to connect direct assessment scores with other data to ensure ecological validity (Wright et al., 2020). Rather than merely making a statement that "these results should be interpreted with caution," it is the responsibility of the practitioner to discuss each of the possible threats to validity.

Crooks, Kane, and Cohen (1996) provide a list of several possible threats to the validity of test scores, including: those internal to the child, those within the environment, and those attributed to the test administrator. Validity threats internal to the child result when test scores reflect more about the child's attitude than their aptitude (e.g., low motivation, test anxiety, and task-related frustration). Validity threats within the environment refer to score differences that occur when tests are given in different modalities, settings, or times (e.g., inappropriate conditions for testing, difficulty providing verbal responses, or difficulty providing nonverbal responses). Finally, validity threats attributed to the test administrator occur when score differences are due to in-consistencies in the test giver's beliefs or actions during testing and interpretation (e.g., point allotment, instructions/demonstrating task expectations, and failure to understand limitations of assessment results). To ameliorate such threats to score validity, the practitioner must bear them in mind during and after testing, then, must describe how these threats were addressed in the written psychological report.

The following provides suggested steps on how to accomplish these tasks.

Steps One and Two: Determining How to Collect the Data

The first step of determining what type of data is needed should be unchanged from pre-COVID-19 practice. As before, data identified as necessary for diagnostic purposes should be based off of apriori questions and case conceptualizations (Reynolds & Kamphaus, 2003; Sattler, 2014, 2018). Figure 2 represents the type of test data required in the color grey (with no border).

Figure 2 provides a representation of how data are collected during the second step of the process. There are three colors (with affiliated outlines) represented in step two. Purple, with a heavy solid outline, represents traditional assessments that were developed to be administered in a computer-based format. Orange, with a dotted outline, represents nontraditional assessments that were developed to be administered in a computer-based format that is not commonly used by school psychologists. And, blue, with a dashed outline, represents tests that were traditionally delivered in a face-to-face format but were adapted to be computer/distance administered. For each option (i.e., purple/solid, orange/dotted, and blue/dot-dash), the authors have provided one or two representative sample tests.
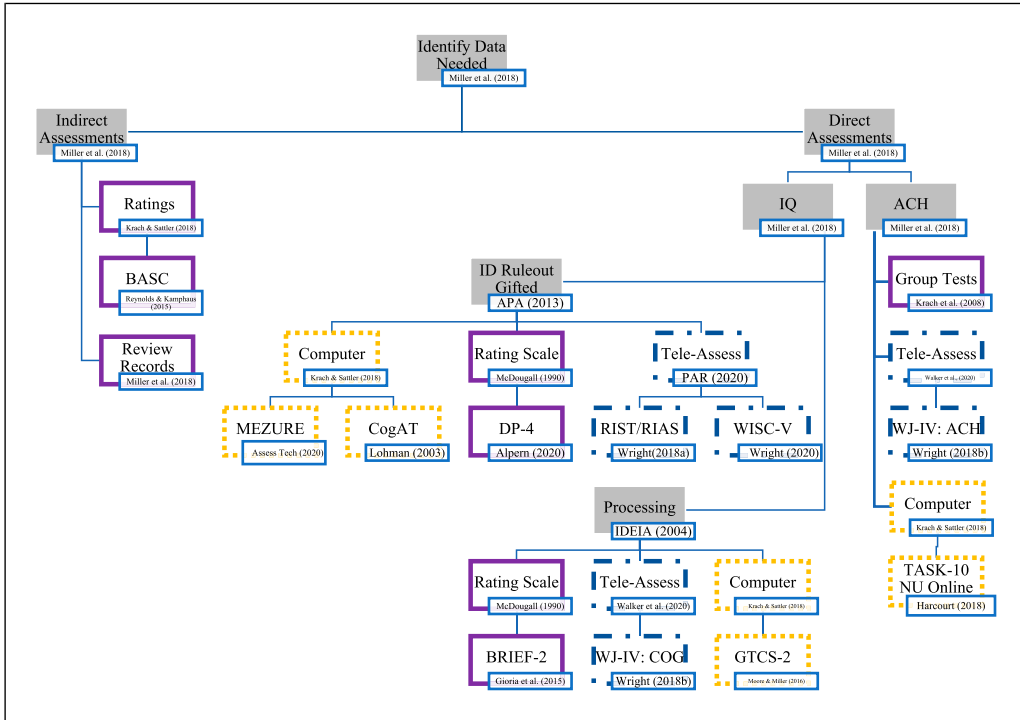
Figure 2. Flowchart for assessment-type selection.
Note. Grey/no outline: task; **purple**/heavy solid outline: traditional computer method; **orange/dotted outline:** uncommon computer method; **blue**/dashed outline: altered to computer method.

In making any decision regarding test selection, it is vital that practitioners undertake only assessments or methods of administration for which they have to meet the ethical competency standards. The APA Principles of Psychology and Code of Conduct (2016) states the following regarding this competency expectation when traditional assessment methods may not be an option:

> 2.01.d
> When psychologists are asked to provide services to individuals for whom appropriate mental health services are not available and for which psychologists have not obtained the competence necessary, psychologists with closely related prior training or experience may provide such services in order to ensure that services are not denied if they make a reasonable effort to obtain the competence required by using relevant research, training, consultation, or study.

In many cases, the school psychology services available during the COVID-19 crisis meet these emergency criteria. Although not explicitly stated in the Code, the implicit expectation is that, as the situation continues, practitioners should receive additional training until they reach competency. Unfortunately, acquiring additional training may be difficult if training opportunities are unavailable due to the novelty of a technique. When that is the case, the Code states that the psychologist is expected to "take reasonable steps to ensure the competence of their work" (2.01.e).

The following are some options available to practitioners who are conducting assessments during the COVID-19 pandemic. Individual psychologists should investigate the instruments in more detail and pursue any required training.

Usual practice/No standardization change (purple – solid line). Instruments in purple/solid were originally developed to be used through tele-assessment methods. In addition, these are instruments that are already frequently used by school psychologists (Sattler, 2014, 2018). Specific, evaluative information about these instruments is readily available through the Buros Institute (Carlson, Geisinger, & Jonson, 2020) as well as through peer-reviewed journals and test manuals. Given the ease of access for other reviews, no further analyses are provided in this study for tests in purple/solid.

Although many of the instruments in purple/solid are considered to be typical practice, COVID-19-related school experiences may introduce psychometric error. For example, many teacher versions of rating scales require that the teacher has personal knowledge of the client for a month or more (Conners, 2008; Reynolds & Kamphaus, 2015a, 2015b). If the child being rated by the teacher has attended school virtually (not in person), then this may violate standardization assumptions of the instrument. Therefore, in those cases, it may be better to depend on self-report ratings (when appropriate), parent report ratings, and/or on historical, classroom performance data. Given common inconsistencies with other sources of data, self-report should not be used as a stand-alone measure to diagnose psychological conditions (Dvorsky, Langberg, Molitor, & Bourchtein, 2016; Möricke, Buitelaar, & Rommelse, 2016; Nelson & Lovett, 2019).

In addition, normative data for social, emotional, and behavioral issues may not be accurate for students amidst a pandemic. A review of psychological effects of long-term, pre-COVID-19, pandemic quarantines found that individuals experienced significantly increased problems with stress and fears (Brooks et al., 2020). Limcaoco, Mateos, Fernandez, and Roncero (2020) found that students, in particular, experienced significantly higher stress levels. Therefore, caution is needed when interpreting tests using pre-COVID-19 norms for standard score conversions related to anxiety, depression, and somatization derived from indirect rating scales. Triangulating social, emotional, and behavioral data using multiple raters, validity indices, direct observations, interviews, and record reviews is more important than ever when making accurate diagnoses.

Unusual practice/No standardization change (orange – dotted line). Instruments in orange/dotted represent those that were developed for computer-based administration; however, these are not as commonly used in psychoeducational, school-based assessments (Sattler, 2018). In addition, many of these may not have been formally evaluated prior to this publication. In general, the following should be considered for any test evaluation (Carlson et al., 2020): (1) What is the theory of the test? (2) How was it developed? and (3) Is the instrument psychometrically sound and normatively appropriate? Table 1 provides specific, example, and evaluative information about several sample instruments.

Usual practice/standardization change (blue – dashes and dots line). Instruments in blue/dot-dash represent tests that have been traditionally administered using a face-to-face method but have some evidence to support their adapted use through tele-assessment. For those with equivalent validity studies, all supporting data should be considered before use in practice. Use of any adapted, tele-assessment instrument lacking concurrent validity data between face-to-face and

Table 1. Equivalency Evaluation: Unusual Practice/No Standardization Change.

| Test | Description | | | Development | | | Technical | | |
	Purpose	Population	Subscales/ composites	Theory	Development	Standardization	Reliability	Validity
CogAT-7	Cognitive Ability Test Learned/ school reasoning test	K-12 grades	VERBAL Picture analogies Sentence completion Picture classification/ verbal classification QUANTITATIVE Number analogies Number puzzles Number series NONVERBAL Figure classification Figure matrices Paper folding OTHER COMPOSITES OVERALL VERBAL + QUANTATIVE QUANTATIVE + NONVERBAL	Vernon, Cattell, and Carroll intelligence and reasoning abilities Builds off research on the reasoning necessary for success in school	Developed from Lorge-Thorndike Intelligence Test (1954) Earlier models only included verbal and nonverbal sections Digital and paper versions	Normed with Iowa assessments 65,630 students K-12 public and private school Some students required some form of accommodation such as extended time or have test read aloud	Coefficient alpha for grades K-12 specific composites range from .8 to .94 Total battery composite ranges from .88 to .97	Correlation between CogAT-7 and the WISC-IV composites range from .68 to .72 Latent factors for the WISC-III and the CogAT-6 correlated .87–Verbal; .67 nonverbal; total composite = .97 (Lohman, 2003a) and or the WJ-III and the CogAT-6, the general factors correlated at .82 (Lohman, 2003b)

(continued)

Table 1. (continued)

Test	Description			Development			Technical	
	Purpose	Population	Subscales/composites	Theory	Development	Standardization	Reliability	Validity
GTCS-2	Cognitive screening test	5–85 years	Short-term memory Long-term memory Visual processing Processing speed Logic & reasoning Auditory processing Word attack ATTENTION GENERAL COGNITIVE ABILITY	CHC theory	Originally, a short, paper-based test Developed with subject matter experts and field testing	2737 total (2014–2016) N = 20 total for ages 5–6 68% white, 13% Black, 3% Asian/PI, 11% Hispanic, <1% Native American, 5% other	Coefficient alpha ages 6–18 ranges from .81 (logic & reasoning) to .97 (visual processing) test–retest child coefficients range from .53 (long-term memory) to .89 (visual processing & word attack)	Correlations between Gibson-2 and WJ-III range from .53 (long-term memory) to .93 (word attack)

Table 1. (continued)

Test	Description			Development			Technical	
	Purpose	Population	Subscales/ composites	Theory	Development	Standardization	Reliability	Validity
Mezure - Children's version	General intelligence test Screening battery Standard battery	Age 6–19	FLUID IQ Visual closure Visual analogies Visual memory Aud. memory CRYSTALLIZED IQ Categorization Information Vocabulary COMPOSITE IQ Other subtests Processing speed Social apperception Aud. memory w. distraction Vis. memory w. distraction	CHC theory	Computer-administered Small pilot studies precede national standardization	Over 5000 subjects Closely matched to US census data (date unknown)	Test–retest coefficients range .64–.92 Split-half (Spearman–Brown) and Cronbach's alpha range .73–.97	Criterion-related validity with WISC-3 range .70–.79; with Iowa test of basic skills range .54–74

(continued)

Table 1. (continued)

Test	Description			Development		Standardization	Technical	
	Purpose	Population	Subscales/ composites	Theory	Development	Standardization	Reliability	Validity
Stanford/ TASK-10 NU	Achievement test	K-12 grades	Reading Mathematics Spelling Language Science Social science Listening	N/A	Originally paper-based Normative update was a combination of paper and online tests No comparison data between versions was reported	360,000 students Stratification variables reflect 2000 census data	KR20 coefficients (internal consistency) range .80s–.90s Alternate forms coefficients range .53–.93 No test–retest reliability reported	Users determine if content matches school curriculum Subtest intercorrelations range .70–.80 Results correlate with OLSAT-8

Note. CogAT-7 = (Cognitive Abilities Test, Form 7; Lohman, 2011); GTCS-2 = (Gibson Test of Cognitive Skills, Second Edition; Moore & Miller, 2016); MEZURE = (Assessment Technologies, Inc., 2020); TASK-10 NU = (Stanford Achievement Test Series, 10th Edition online with normative update; Harcourt Assessment Inc., 2018); CHC = (Cattell–Horn–Carroll; Schneider & McGrew, 2012).

tele-assessment versions is not recommended. To feel truly comfortable with version equivalency, all of these standards should be met:

1. research should be published in a reputable, academic source;
2. between version correlations should be at or above .80 (Krach, McCreery et al., 2020). Concurrent validity standards are not sufficient for equivalency studies;
3. no statistically significant differences between scores derived via different formats (AERA, APA, & NCME, 2014; APA, 1986);
4. between version effect sizes should be at or below .2 (Daniel, Wahstrom, & Zang, 2014);
5. normative dispersion of scores should have a statistically similar shape (AERA et al., 2014; APA, 1986);
6. samples used in the study should match the U.S. Census (2019) for race/ethnicity and the U.S. Department of Education (2019) for disability status (Grosch, Gottlieb, & Cullum, 2011; Hodge et al., 2019; Krach, McCreery et al., 2020);
7. any sample used in the study should have a large enough size to be able to run the statistical analyses used in the study (Faul et al., 2007; Hulley, Cummings, Browner, Grady, & Newma, 2013; Lakens, 2017).

Specific instruments. The more assessment error that is introduced, the more vital it becomes for practitioners to use multiple methods of collecting data. Error can be estimated for each method based on consideration of the deviation from standardized and professional practice. The following section provides evaluation guidelines and examples of tests that may be considered from the orange/dotted and the blue/dot-dash groups. The specific tests provided for those options can be found in Figure 2 and are also evaluated in Tables 1–3.

Orange/dotted: Unusual practice/no standardization change. Table 1 provides an analysis of the three instruments described in the flowchart as "unusual practice/no standardization change." All three tests were originally validated to be used as computer-administered assessments. The columns in Table 1 were selected as they match the main areas evaluated in the Buros Organization for Test Reviews (n.d.). These three instruments are merely examples. There are other, computer-based instruments available. Each additional instrument would need to be evaluated in a similar manner by the user prior to administration.

For the Cognitive Abilities Test, Form 7 (CogAT-7; Lohman, 2011), a Buros review is available (Ackerman & Miller, 2017). This assessment is used to measure cognitive ability as well as general reasoning. Overall, both reviewers describe the test positively. The only problematic issues described were concerns that much of the composite validity findings are based on previous versions of the tests. In addition, there is a concern about the lack of psychometric data provided for the subscales. One advantage listed when using this test is that it provides more pictorial options, making it fairer when used with English language learners.

No Buros review has been published for the Gibson Test of Cognitive Skills, Second Edition (GTCS-2; Moore & Miller, 2016). The GTCS was originally designed as a screener, but the authors expanded it to provide a comprehensive composite when the second edition was published. Given that the technical manual is available online at the publisher's website, the information needed for a personal review of the test is readily available. From the information available, the reliability and validity data seem to support the use of the composites derived from this instrument. The subscales, especially the long-term memory subscale, should be interpreted with more caution. One additional advantage of this test is that it is available in multiple languages.

Finally, no Buros review has been published for the MEZURE (Assessment Technologies, Inc, 2020). The MEZURE provides interactive computer-administered assessments of general intelligence, distractibility, and emotional apperception with two standardized versions (child and adult). Based on information from the clinical manual, the reliability data are acceptable. The

Table 2. Equivalency Evaluation: Usual Practice/Standardization Change.

Test name	Study citation	In-person and tele-assessment	Reported r	Statistical significance (effect size) Sig	ES	Score dispersion shape comparison	Sample demographics	Sample size (power analysis) Study n	Needed Stat	Needed n
	In a peer-reviewed journal	Subtests and composite scores should be evaluated separately	≥ .80	Desired not sig for t-test	Desired ≤ .2	Desired not significantly different	Does it match the U.S. Census (2019)? [76.3% white, 13.4% Black, 18.5% Hispanic, 5.9% Asian, 1.3% Native American, 2.8% mixed] Does it match the U.S. DOE (2018)? [13.7% have children with disabilities] Does it match your client?		Matched pairs t-test / CI / r	156 pairs (312) / 176 pairs (352) / 10
WJ-IV: COG	Wright (2018b)	CogAT-6 General intellectual ability	None	.226	−.159	None reported	Matched pairs between age 5 and 16 with 120 men and 120 women in the study. Of traditional admin, 48.3% white, 15.8% Black, 31% Latino, 4.1% Asian, & .8% Native American. The online admin had 68% white, 4.1% Black, 20/5% Latino, 3.2% Asian, & 4.1% Native American	240	T-test	156 (pairs) or 312 (total)
		Gf-Gc composite		.485	.090					
		Comp-knowledge		.747	−.042					
		Fluid reasoning		.211	.162					
		Short-term working memory		.606	.067					
		Cognitive efficiency		.139	.192					
		Oral vocabulary		.968	−.006					
		Number series		.474	.093					
		Verbal attention		.351	−.120					
		Letter-pattern matching		.390	.111					
		Phonological processing		.744	.042					
		Story recall		.122	.194					
		Visualization		.476	.093					
		General information		.638	−.061					
		Concept formation		.181	.173					
		Numbers reversed		.137	.192					

Table 2. (continued)

Test name	Study citation	In-person and tele-assessment	Reported r	Statistical significance (effect size)		Score dispersion shape comparison	Sample demographics	Sample size (power analysis)		
				Sig	ES			Study n	Needed Stat	Needed n
WJ-IV: COG	Unclear	Met	Not reported	Met	Met	Not reported	In person does not match the census			Met

Notes 1. WJ-IV: COG = Woodcock–Johnson Tests of Cognitive Ability, Fourth Edition; WJ-IV: ACH = Woodcock–Johnson Tests of Achievement, Fourth Edition; RIASs = Reynolds Intellectual Assessment Scales; TOGRA = Test of General Reasoning Ability; RAIT = Reynolds Adaptable Intelligence Test; WISC-V = Wechsler Intelligence Scales for Children, Fifth Edition; CogAT = Cognitive Abilities Test.
Note 2. Correlation requirement ≥ .80 established by Krach, McCreery et al. (2020).
Note 3. Effect size of less than .2 established by Daniel et al. (2014).
Note 4. Sample equivalency guidelines provided by Grosch et al. (2011), Hodge et al. (2019) and Krach, McCreery et al. (2020).
Note 5. Data on population breakdown derived from the U.S. Census Bureau (2019) and US-DOE and NCES (2019).
Note 6. Power analysis using G*Power (Faul et al., 2007) for a matched paired t-test based on α = .05, 1−β = .8, and effect size = .2 (Daniel et al., 2014) requires n = 156; power analysis using Statulator with same criteria n = 199 (where n = number of pairs; Dhand & Khatkar, 2014).
Note 7. Correlation minimal sample size derived from Hulley et al. (2013).
Note 8. Confidence interval comparison sample size derived from Lakens (2017).
Note 9. Statistical significance, correlation requirements, and normative shape distribution requirements were established by APA (1986) and AERA et al. (2014).

Table 3. Additional Direct-Assessment and Tele-Assessment Equivalency Studies.

Test name	Study citation	In-person and tele-assessment	Reported r	Statistical significance (effect size)		Score dispersion shape comparison	Sample demographics	Sample size (power analysis)		
				Sig	ES			Study n	Needed Stat	n
RIAS	Wright (2018a)	Guess what	None	.203	.006	None reported	Matched pairs of individuals between ages 3 and 19 with 52 men and 52 women. Of the traditional admin, 63% white, 12% Black, 19% Hispanic, and 6% other. The online admin was similar with 62% white, 12% Black, 21% Hispanic, & 6% other	104	t-test	156 (pairs) or 312 (total)
		Odd-item out		.770	−.009					
		Verbal reasoning		.373	−.002					
		What is missing		.089	.018					
		Verbal memory		.942	−.010					
		Nonverbal memory		.080	.020					
		Speeded naming task		.063	.024					
		Speeded picture search		.094	.018					
		Verbal Intelligence Index		.258	.003					
		Nonverbal Intelligence Index		.185	.007					
		Composite Intelligence Index		.155	.010					
		Composite Memory Index		.414	−.003					
		Speeded Processing Index		.034	.034					
TOGRA/ RAIT	Manual; Reynolds and PAR (2020)	None reported	None	None	None	None	None	?	N/A	?

Table 3. (continued)

Test name	Study citation	In-person and tele-assessment	Reported r	Statistical significance (effect size)		Score dispersion shape comparison	Sample demographics	Sample size (power analysis)		
				Sig	ES			Study n	Needed Stat	n
WISC-V	Wright (2020)	None reported Authors provided 2 separate independent intercorrelation matrices for subtests and indices: one for online administration and one for in person	None	Letter-number sequencing found to be "significantly higher ... in person." (no values reported)	None reported	None reported	Matched pairs between ages 6 and 16 with 133 men and 123 women. Pairs were matched on age, gender, and K-BIT 2 scores. Demographics only provide parent's education level (83% had some college)	256	CI	176 (pairs) or 352 (total)
WJ-IV: ACH	Wright (2018b)	Broad reading	None	.725	−.045	None reported	Matched pairs between ages 5 and 16 with 120 men and 120 women. Traditional admin, 48.3% white, 15.8% Black, 31% Latino, 4.1% Asian, and .8% Native American. The online admin had 68% white, 4.1% Black, 20.5% Latino, 3.2% Asian, & 4.1% Native American	240	T-test	156 (pairs) or 312 (total)
		Broad mathematics		.302	−.134					
		Broad writing		.219	−.159					
		Letter-Word Identification		.544	−.079					
		Applied problems		.460	−.096					
		Spelling		.204	−.165					
		Passage comprehension		.992	.001					
		Calculation		.488	−.090					
		Writing samples		.592	−.069					
		Word attack		.715	.047					
		Oral reading		.452	−.098					
		Sentence reading fluency		.626	−.063					
		Math facts fluency		.432	−.102					
		Sentence writing fluency		.285	−.139					

Note. RIAS = Reynolds Intellectual Assessment Scales; TOGRA = Test of General Reasoning Ability; RAIT = Reynolds Adaptable Intelligence Test; WISC-V = Wechsler Intelligence Scales for Children; WJ-IV: ACH = Woodcock–Johnson Tests of Achievement, Fourth Edition.

validity data suggest the MEZURE is appropriate for use with various populations; however, criterion-related validity was established using an outdated test (*Wechsler Intelligence Scales for Children, Third Edition* [WISC-III]), and it was not clear from the manual in what year this instrument was developed and normed. However, standardization occurred in all 50 states, with efforts to include traditionally underrepresented groups (e.g., Eskimos/Aleut Islanders and Native American Indians) in the sample. An advantage of the MEZURE is that it is available in Spanish and Russian, in addition to standard American English.

Blue/dot-dash: Usual practice/standardization change. Error can be introduced if testing methods differ from the standardization methods used during the normative data collection (Sattler, 2018). In the case of COVID-19, the difference is usually found in the practice of moving from a face-to-face to a tele-assessment method. Adaptation equivalency guidelines were set forward by the American Psychological Association (APA, 1986) in the *Guidelines for Computer-Based Tests and Interpretations* as well as by a joint commission in the *Standards for Educational and Psychological Testing* (AERA et al., 2014). Additional equivalency requirements were set forward by Grosch et al. (2011), Hodge et al. (2019), Krach, McCreery et al. (2020), Cohen (1988), and Farmer et al. (2020).

Table 2 lists these guidelines as well as the parameters for how to evaluate each one when reading an equivalency study. Ideally, if the two versions are equivalent, all of the requirements should be listed as "met." The further away from this standard, the less equivalency can be assured. Table 3 provides a visual representation of this information for one such study. The following is a descriptive example evaluation from the same study.

An equivalency study on the *Woodcock–Johnson Tests of Cognitive Ability, Fourth Edition* (WJ-IV: COG; Wright, 2018b) was analyzed at the bottom of Table 2 as an example. The Wright (2018b) article was published in a journal sponsored by the American Board of Assessment Psychology. Upon review of the journal, there was no evidence of an impact factor or that the journal is peer-reviewed. The study examined subscales and composites separately using matched pair, t-tests with an inappropriately small sample size. None of the t-tests were significant; however, this may have been due to the small sample size or the small effect sizes. The sample used did not match the census for the face-to-face version, but the match improved for the online administration. Finally, no correlations or score dispersion data were provided. Overall, this one study of the WJ-IV meets some, but not all, of the equivalency requirements. Table 3 provides the necessary data on a few other equivalency studies for other assessments.

Case Study: Example Use of Guide

The following is an example of how this guide can be used with a client. Mock information is provided for this case study. Thomas is the oldest of three children from an immigrant family from Mexico. He is third-generation to the United States; his parents speak English as the primary language at home. Both of his parents work full-time during school hours. Thomas's school shut down due to COVID-19; he is in sixth grade (age 12) and cares for his siblings (first and third grade) while his parents are at work. School is now taught completely online for all three children. There is one computer in the home; it is an older model. The family has reliable internet service.

Prior to COVID-19, his school provided response to intervention (RTI) services at the Tier 1 and Tier 2 level for Thomas in the area of mathematics. The child study team determined that Thomas failed to respond to the intervention. His school district requires several additional criteria beyond failure to respond to an intervention before making an SLD diagnosis. First, they require that cognitive processing deficits be identified as well as comprehensive, cognitive data to rule out an intellectual disability. In addition, they require achievement data showing that he is performing below expectations when compared to other students his age. Finally, all additional exclusionary clause factors must be evaluated to rule out nondiagnostic factors as the primary cause of his

problem (i.e., emotional, cultural, environmental, economic, sensorimotor, medical/neurological issues, or lack of educational opportunity).

Step One: Determine the Type of Data Needed

1. General cognitive ability ("g") for ID rule out.
   a. Assessment of general intellectual ability (GIA).
2. Cognitive processes related to mathematics achievement (Geary, 2011).
   a. Assessment of processing speed.
   b. Assessment of working memory.
3. Math ability.
   a. Math reasoning and math calculation assessment.
   b. Math reasoning and math calculation review of records.
4. Emotional issues rule outs.
   a. Rating scales for anxiety, depression, and psychosis.
   b. Interviews (child, parent, and teacher discussing current and pre-COVID-19 status).
5. Other exclusionary factor rule-outs.
   a. Review of records.
   b. Vision/hearing screening.
   c. Review of RTI data.
   d. Parent and teacher interviews/questionnaires.

Step Two: How Will the Data be Collected

Qualitative. All parent and teacher interviews will be collected using a telephone or video teleconferencing method (e.g., Teams, Zoom, etc.). Discussions with his parents and his previous teachers may have to take place in the evenings or on the weekends. All parent and teacher questionnaires can be sent through email or through the US Postal Service (USPS) with self-addressed stamped envelopes (SASEs) included. All school record reviews (including vision/hearing results) can be accessed from the school.

Quantitative – rating scales. An emotional rating scale, such as the *Behavior Assessment System for Children, Third Edition* (BASC-3; Reynolds & Kamphaus, 2015a, 2015b), can be completed through an online link emailed to the rater or can be sent through the USPS with a SASE. Raters could include Thomas's parents, his current teachers, his pre-COVID-19 teachers, and Thomas.

Quantitative – rating scales. In addition to social, emotional, and behavioral issues, the BASC-3 also assesses executive functioning skills including problem solving skills. This information may be useful when exploring underlying cognitive processing deficits. The *Brown Executive Function/Attention Scales* (Brown, 2019) rating scale provides working memory information. Both the BASC-3 and the Brown EF/A have been evaluated by the Buros Institute and would fall in the "purple/solid" category of the guide.

Quantitative – cognitive. A test from the orange/dotted category that might be applicable for Thomas would be the GTCS-2 (Moore & Miller, 2016). The GTCS-2 provides a general cognitive ability composite score as well as subscale scores for short-term memory, long-term memory, and processing speed. Although the GTCS-2 is not a traditional test given by school psychologists, the psychometrics are sound with a large enough sample size that includes a representative sample of Hispanic students.

As a backup assessment of his general cognitive ability (or "g"), a test from the blue/dot-dash category could be chosen. In this case, the Gf-Gc composite score of the WJ-IV: COG could be used. The WJ-IV: COG could not be used as the primary measure of "g" because the single equivalency study available failed to meet several of the guidelines listed in Table 3 (Wright, 2018b). The Gf-Gc would be chosen over the GIA as the better measure of "g" for an adapted tele-assessment version of the WJ-IV: COG because the GIA is comprised of seven subtests; one of which has an effect size of .19 (.20 is the cut-off). In comparison, the Gf-Gc Composite alternative is comprised of only four subtests, none of which have effect sizes close to the recommended cutoff. Neither of the cognitive processing composites for the WJ-IV: COG in memory or speed would be recommended. The cognitive processing speed composite is not available as one of the subscales required was not included in the equivalency research study. In addition, one of the two subscales in the short-term working memory composite had an effect size of .19, which is very close to the cutoff.

Unfortunately, none of the other measures in blue/dot-dash, including the Reynolds Intellectual Assessment Scales, Second Edition (Reynolds & Kamphaus, 2015a, 2015b) and the WISC-V, could be used as a primary measure for Thomas because their related studies also did not provide enough data to demonstrate equivalency (Wright, 2018a, 2020).

Quantitative − achievement. For direct assessment achievement data, the tele-assessment administration of the *Woodcock−Johnson Tests of Achievement, Fourth Edition* (WJ-IV: ACH) would be a possible option for Thomas. The applied problems, calculation, and math facts fluency subscales were not statistically or practically different across administration methods (Wright, 2018b). In addition, the sample in Wright's study was sufficient in Latinx representation. Unfortunately, as with the WJ-IV: COG, the sample size for the ACH is too small to ensure that the tele-assessment version is equivalent. Therefore, as with the COG, ACH test results may be helpful only when combined with other achievement data. Specifically, the information provided should be compared with his grades in mathematics, his previous scores on group administered mathematics tests, and teacher/parent reports on his ability.

Step Three − Practical Issues

Access is a considerable concern when choosing testing methods. In the case of Thomas, the first question would be: does his family have sufficient computer access to be able to do the assessments chosen? The answer to that question is that it depends on the tests being given. Access should not be an issue with gathering qualitative data and rating scales due to their low-tech nature. However, because the GTCS-2 is computer-based, access to the technology needed for this test may be more complicated. Thomas must be able to take the test in a quiet environment. He must be proctored to verify that he is the one taking the test. Proctoring can occur through the use of a signed parent affidavit or through the use of an outside cell phone camera facing him while he completes the testing session. In general, cell phones are considered to be HIPPA-compliant when used in this manner (CPH & Associates, n.d.).

The WJ-IV: ACH and COG subscales can be administered through PresenceLearning's online tools (https://www.presencelearning.com). Given that WJ-IV manipulatives are limited to a response booklet, Thomas is old enough to start and end the session without the need of an additional support person. The Presence Learning online tools do not need an additional test proctor.

Before starting any of tests, the test administrator must establish all of the following. The family must: (1) have a quiet environment with few distractions with a computer available, (2) have access to a phone with a camera, (3) have the skills needed to implement the testing procedures required, and (4) be provided training on preparing the test-taking technology and the environment. Finally, before starting any testing, it is vital that appropriate consent be provided by his

parents. This consent would need to include traditional consenting information as well as any deviations from standard practice.

Step Four – Interpretation

Once the assessment is complete, it is time to interpret the findings. When interpreting the findings for Thomas during COVID-19, several issues must be considered. First, any standardization differences must be considered and discussed. These include the use of computer-mediated assessments, but these also include any distractions in the room, any issues related to how materials were provided and returned, and the need for any additional people to aid with the test environment. When interpreting social, emotional, and behavioral results, do not forget to consider any differences between Thomas's functioning before and during COVID-19. Additional information should be included specific to new stressors in Thomas' life. In addition, consideration is needed regarding any environmental issues that might be hindering his learning. For example, if Thomas is in charge of his siblings, as well as his own learning, how might that affect his overall academic achievement performance? All of these issues must be discussed and considered in both verbal and written interpretations.

In the case of Thomas, there is a clear need for multiple sources of data. Qualitative data are used to identify problem patterns in potential systemic and diagnostic issues. Quantitative data triangulate results from a combination of computer-based assessments, adapted tele-assessments, and rating scales. Both quantitative and the qualitative data must converge until a singular diagnostic pattern emerges. Although that is true in most cases, it is especially true during the COVID-19 situation, given the sources of potential error introduced in this data collection model.

Limitations/Call for Future Discussion

There are two main limitations of the proposed guide. The first is that only one or two tests were evaluated for each criterion. Although more traditional tests could be considered in this article, there was no room in the current study to address them all. More instruments in the categories of computer-based and adapted tele-assessment were not addressed because of limited published research.

The second limitation is of greater concern: the guide may not be applicable to a Multi-Tiered System of Support (MTSS) model. This is because the interwoven nature of data collection and interventions in MTSS makes decision-making more complicated than the options provided in Figures 1 and 2. For MTSS, the practitioner would need to evaluate validity for practice differences in interventions as well as assessments. For example, if the intervention had been empirically studied only for use in a face-to-face setting, it may not be valid to use it in a virtual one. This deviation from established intervention practice only becomes compounded when added to the assessment deviations listed earlier in this study. In addition, if the assessment requires tracking observable behaviors in situ, then it may be difficult to get accurate data in a virtual environment. Other assessment methods may need to be considered. Given all of these factors, a model for MTSS diagnoses would need a guide of its own.

Conclusion

COVID-19 has changed the way school-based professionals interact with their students. However, there has been no associated change in the child find mandate of federal law (IDEIA, 2004). Eligibility decision deadlines have not been put on hold. Therefore, practitioners need a guide for deciding which assessments to give when face-to-face testing is restricted. This study provided

a proposed set of steps for making assessment decisions based on data available about a variety of types and modalities of assessment tools.

Examples of possible tests were discussed throughout the study. Current research for adapted tele-assessment methods was limited to a single, problematic study for each example. Unfortunately, none of the studies for these adapted tele-assessments met all of the equivalency requirements to ensure acceptable substitution of traditional face-to-face methods. Therefore, any tests listed here as usual practice/standardization change (blue/dot-dash) should, for now, be considered only as support for other test data obtained in some other manner. This may change as additional studies on these instruments come to light.

The goal of this study has been to provide practitioners with one possible process that will assist them in meeting child find mandates to identify disabilities when present, while still maintaining ethical standards for assessment in the wake of COVID-19 disruptions to schooling. Specifically, the authors described traditional, optional, and alternative methods for gathering assessment data. As school psychologists are faced with assessment backlogs and new referrals, this guide may allow for traditional and nontraditional assessment methods to be used together to make eligibility decisions, while maintaining legal standards and ethical best practices in the areas of test selection, administration, and interpretation.

Declaration of Conflicting Interests

Funding

ORCID iD

S. Kathleen Krach ![ORCID] https://orcid.org/0000-0002-6853-379X

References

Ackerman, P. L., & Miller, D. M. (2017). Test review of the cognitive abilities test, form 7. In J. F. Carlson, K. F. Geisinger, & J. L. Jonson (Eds.), *The twentieth mental measures yearbook.* Lincoln, NE: Buros Center for Testing.

American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME]. (2014). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

American Psychological Association. (2017). *Ethical principles of psychologists and code of conduct.* Washington, DC: Author.

American Psychological Association Committee on Professional Standards and Committee on Psychological Tests and Assessment. (1986). *Guidelines for computer-based tests and interpretations.* Washington, DC: Author.

Assessment Technologies, Inc. (2020). *MEZURE clinical manual.* New York, NY: Author.

Bramlett, R. K., Murphy, J. J., Johnson, J., Wallingsford, L., & Hall, J. D. (2002). Contemporary practices in school psychology: A national survey of roles and referral problems. *Psychology in the Schools, 39*(3), 327-335. doi:10.1002/pits.10022

Brooks, S. K., Webster, R. K., Smith, L. E., Woodland, L., Wessely, S., Greenberg, N., & Rubin, G. J. (2020). The psychological impact of quarantine and how to reduce it: Rapid review of the evidence. *Lancet, 395*(10227), 912–920. doi:10.1016/S0140-6736(20)30460-8

Brown, T. E. (2019). *Brown executive function/attention scales.* Bloomington, MN: Pearson.

Buros Center for Testing. (n.d.). *Organization of test reviews for the mental measurements yearbook Series.* Retrieved from https://buros.org/organization-test-reviews-mental-measurements-yearbook-series

Carlson, J. F., Geisinger, K. F., & Jonson, J. L. (Eds.). (2020). *The twenty-first mental measurements yearbook.* Lincoln, NE: Buros Center for Testing.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Earlbaum Associates.

Conners, C. K. (2008). *Conners 3.* North Tonawanda, NJ: MHS.

C.P.H. & Associates. (n.d.). *How to be HIPPA compliant with your mobile device.* Retrieved from https://www.cphins.com/how-to-be-hipaa-compliant-with-your-mobile-device/

Crooks, T. J., Kane, M. T., & Cohen, A. S. (1996). Threats to the valid use of assessments. *Assessment in Education: Principles, Policy & Practice, 3*(3), 265-286. doi:10.1080/0969594960030302

Daniel, M. H., Wahstrom, D, & Zang, O. (2014). Q–interactive equivalence of Q-interactive® and paper administrations of cognitive tasks: WISC®–V. (Q-Interactive Technical Report 8). Pearson.

Dhand, N. K., & Khatkar, M. S. (2014). *Statulator: An online statistical calculator. Sample size calculator for comparing two paired means.* Retrieved from http://statulator.com/SampleSize/ss2PM.html

Dvorsky, M. R., Langberg, J. M., Molitor, S. J., & Bourchtein, E. (2016). Clinical utility and predictive validity of parent and college student symptom ratings in predicting an ADHD diagnosis. *Journal of Clinical Psychology, 72*(4), 401-418. doi:10.1002/jclp.22268

Farmer, R. L., McGill, R. J., Dombrowski, S. C., Benson, N. F., Smith-Kellen, S., Lockwood, A. B., Powell, S., … Stinnett, T. A. (2020). Conducting psychoeducational assessments during the COVID-19 crisis: The danger of good intentions. *Contemporary School Psychology.* Advance online publication. doi:10.1007/s40688-020-00293-x

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods, 41*(4), 1149-1160. doi:10.3758/BRM.41.4.1149

Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology, 47*(6), 1539-1552. doi:10.1037/a0025510

Grosch, M. C., Gottlieb, M. C., & Cullum, C. M. (2011). Initial practice recommendations for tele-neuropsychology. *The Clinical Neuropsychologist, 25*(7), 1119-1133. doi:10.1080/13854046.2011.609840

Harcourt Assessment, Inc. (2018). *Stanford achievement test Series, tenth edition online with normative update (TASK-10 NU).* San Antonio, TX: Pearson.

Henke, J. (2020, March 30). *Webinar: Guidelines for administration via telepractice.* Bloomington, MN: Pearson.

Hodge, M. A., Sutherland, R., Jeng, K., Bale, G., Batta, P., Cambridge, A., Detheridge, J., , … Silove, N. (2019). Agreement between telehealth and face-to-face assessment of intellectual ability in children with specific learning disorder. *Journal of Telemedicine and Telecare, 25*(7), 431-437. doi:10.1177/1357633X18776095

Hulley, S. B., Cummings, S. R., Browner, W. S., Grady, D., & Newman, T. B. (2013). *Designing clinical research: An epidemiologic approach* (4th ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

Individuals with Disabilities Education Improvement Act, 20 U.S.C. § 1400. (2004).

Krach, S. K., McCreery, M. P., Dennis, L., Guerard, J., & Harris, E. L. (2020). Independent evaluation of Q-Interactive: A paper equivalency comparison using the PPVT-4 with preschoolers. *Psychology in the Schools, 57*(1), 17-30. doi:10.1002/pits.22325

Krach, S. K., Paskiewicz, T. L., & Monk, M. M. (2020). Testing our children when the world shuts down: Analyzing recommendations for adapted tele-assessment during COVID-19. *Journal of Psycho-educational Assessment,* online. doi:10.1177/0734282920962839

Krach, S. K., & Sattler, J. M. (2018). Computer-based administration, scoring, and report writing. In J. M. Sattler (Ed.), *Assessment of children: Cognitive foundations* (6th ed., pp. 47-50). La Mesa, CA: Jerome M. Sattler, Publisher, Inc.

Lakens, D. (2017). Equivalence tests: A practical primer for *t* tests, correlations, and meta-analyses. *Social Psychological & Personality Science, 8,* 355-362. doi:10.1177/1948550617697177

Limcaoco, R. S. G., Mateos, E. M., Fernandez, J. M., & Roncero, C. (2020). *Anxiety, worry and perceived stress in the world due to the COVID-19 pandemic, March 2020. Preliminary results*. MedRxiv. doi:10.1101/2020.04.03.20043992

Lohman, D. F. (2003a). *The Wechsler intelligence scale for children III and the cognitive abilities test (Form 6): Are the general factors the same?* Retrieved from https://pdfs.semanticscholar.org/1730/b8339f3f5472596eff694a6f543bc34f3252.pdf

Lohman, D. F. (2003b). *The Woodcock-Johnson III and the cognitive abilities test (Form 6): A concurrent validity study.* Retrieved from https://pdfs.semanticscholar.org/bd2e/ed7e216abd6ffa9d71b15d2920583ed955d1.pdf

Lohman, D. F. (2011). *Cognitive Abilities test, Form 7.* Rolling Meadows, IL: Riverside Publishing.

Miller, F. G., Johnson, A. H., Yu, H., Chafouleas, S. M., McCoach, D. B., Riley-Tillman, T. C., … Welsh, M. E. (2018). Methods matter: A multi-trait multi-method analysis of student behavior. *Journal of School Psychology, 68*, 53-72. doi:10.1016/j.jsp.2018.01.002

Moore, A. L., & Miller, T. (2016). *Gibson test of cognitive skills – version 2: Digital and interactive test. Technical manual for centers*. Gibson Institute of Cognitive Research. Retrieved from https://www.thegibsontest.com/wp-content/uploads/2019/06/GT-V2-Digital-Interactive-Technical-June-2019.pdf

Möricke, E., Buitelaar, J. K., & Rommelse, N. N. (2016). Do we need multiple informants when assessing autistic traits? The degree of report bias on offspring, self, and spouse ratings. *Journal of Autism and Developmental Disorders, 46*(1), 164-175. doi:10.1007/s10803-015-2562-y

National Association of School Psychologists. (2020a). *Virtual service delivery in response to COVID-19 disruptions*. Retrieved from https://www.nasponline.org/resources-and-publications/resources-and-podcasts/school-climate-safety-and-crisis/health-crisis-resources/virtual-service-delivery-in-response-to-covid-19-disruptions

National Association of School Psychologists. (2020b). *Telehealth: Virtual service delivery updated recommendations.* Retrieved from https://www.nasponline.org/resources-and-publications/resources-and-podcasts/covid-19-resource-center/special-education-resources/telehealth-virtual-service-delivery-updated-recommendations

Nelson, J. M., & Lovett, B. J. (2019). Assessing ADHD in college students: Integrating multiple evidence sources with symptom and performance validity data. *Psychological Assessment, 31*(6), 793-804. doi:10.1037/pas000070

PAR. (2020). *Partnering with you: Using telehealth for psychological assessment.* Retrieved from https://www.parinc.com/Using-PAR-digital-assessments-during-the-COVID-19-crisis

Reynolds, C. R., & Kamphaus, R. W. (Eds.). (2003). *Handbook of psychological and educational assessment of children: Personality, behavior, and context* (2nd Ed.). New York: Guilford Press.

Reynolds, C. R., & Kamphaus, R. W. (2015a). *BASC-3 behavioral and emotional screening system manual.* Lutz, FL: Pearson.

Reynolds, C. R., & Kamphaus, R. W. (2015b). *Reynolds intellectual assessment scales* (2nd ed.). Lutz, FL: PAR.

Reynolds, C. R., & PAR Staff. (2020). *Remote administration: RAIT and TOGRA*. PAR. Retrieved from https://www.parinc.com/Portals/0/Webuploads/samplerpts/RAIT_TOGRA_Digital%20Paper_FINAL.pdf

Sattler, J. M. (2014). *Foundations of behavioral, social and clinical assessment of children.* (6th ed.). La Mesa, CA: Jerome M. Sattler, Publisher, Inc.

Sattler, J. M. (2018). *Assessment of children: Cognitive foundations and applications* (6th ed.). La Mesa, CA: Jerome M. Sattler, Publisher, Inc.

Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 99-144). The Guilford Press.

United States Census Bureau. (2019, July 1). *Quick Facts: United States*. Retrieved from https://www.census.gov/quickfacts/fact/table/US/PST045219

United States Department of Education. (2019). *Individuals with Disabilities Education Act (IDEA) database*. Office of Special Education Programs. Retrieved from https://www2.ed.gov/programs/osepidea/618-data/state-level-data-files/index.html#bcc

United States Department of Education. (2020, March). *Fact sheet: Addressing the risk of COVID-19 in schools while protecting the civil rights of students*. Retrieved from https://www2.ed.gov/about/offices/list/ocr/docs/ocr-coronavirus-fact-sheet.pdf

United States Department of Education, & National Center for Education Statistics. (2019). *Digest of education statistics, 2018 (NCES 2020-009)*. Retrieved from https://nces.ed.gov/fastfacts/display.asp?id=64

United States Social Security Administration [SSA]. (2020). *Social security and Coronavirus disease (COVID-19)*. Retrieved from https://www.ssa.gov/coronavirus/

Walker, K. E., Taylor, S., Wright, J., & Johnsrud, N. (2020a, May 5). *Tele-Assessment: What you Need to know [webinar]*. Presence Learning in Coordination with Riverside Publishing. Retrieved from https://www.presencelearning.com/sped-ahead-webinar/tele-assessment-what-you-need-to-know/

Walker, K. E., Taylor, S., Wright, J., & Johnsrud, N. (2020b, May 5). *Online-SPED services and assessment [Webinar]*. Presence Learning in Coordination with Riverside Publishing.

Wright, A. J. (2018a). *Equivalence of remote, online administration and traditional, face-to-face administration of the Reynolds Intellectual Assessment Scales-Second edition (Online white paper)*. Presence Learning.

Wright, A. J. (2018b). Equivalence of remote, online administration and traditional, face-to-face administration of Woodcock- Johnson IV cognitive and achievement tests. *Archives of Assessment Psychology*, *8*(1), 23-35.

Wright, A. J. (2020). Equivalence of remote, digital administration and traditional, in-person administration of the Wechsler Intelligence Scale for Children, Fifth Edition (WISC-V). *Psychological Assessment*. Advance online publication. doi:10.1037/pas0000939

Wright, A. J., Mihura, J. L., Pade, H., & McCord, D. M. (2020, April).*Guidance on psychological tele-assessment during the COVID-19 crisis*. American Psychological Association. Retrieved from https://www.apaservices.org/practice/reimbursement/health-codes/testing/tele-assessment-COVID-19