One

THE LANDSCAPE OF PSYCHOLOGICAL TELE-ASSESSMENT

The COVID-19 pandemic has changed the landscape of health and mental healthcare across the globe. And it seems that it has changed the landscape forever. While the psychotherapy functions of psychologists and other mental health professionals have relatively easily transitioned to an online modality, one which has had significant empirical inquiry and support (Batastini, King, Morgan, & McDaniel, 2016; Bolton & Dorstyn, 2015; Reese, Slone, Soares, & Sprang, 2015; Varker, Brand, Ward, Terhaag, & Phelps, 2019), the psychological assessment functions of psychologists have had to rapidly adapt within a context of significantly less empirical support. The requirements for some, especially performance-based, tests—using manipulatives and physical stimuli—have placed a disproportionate burden on psychological assessment to figure out ways to adapt to a remote, online environment.

Many clinicians chose to pause their assessment services, expecting the world to resume its in-person engagement capabilities relatively soon. However, it is becoming clearer that while a vaccine will likely emerge at some point, COVID-19 may permanently alter health and mental health practices. Thus, clinicians need to consider how they can adapt and continue to deliver essential assessment services to clients, whether in a tele-assessment or somehow physically distanced format.

While psychological tele-assessment has certainly garnered a great deal of attention lately, some researchers have been working toward an evidence base for it for many years. The impetus for such research agendas seems originally to be related more to equity and access, ensuring that those in remote or rural areas have access to needed assessments. This reasoning also accounts for why so much of the research to date evaluating tele-assessment has focused on older adults,

Essentials of Psychological Tele-Assessment, First Edition.
A. Jordan Wright and Susan Engi Raiford

neuropsychological assessment, and specifically assessment of cognition (e.g., dementia evaluations). While it seems somewhat counterintuitive that this population would naturally gravitate toward technology, these older adults are also more likely to be home-bound and have difficulty getting to appointments (Qiu et al., 2010). Because of these lines of research, we are lucky to at least have some empirical evidence base for remote, online tele-assessments (in contrast to assessments conducted with social distancing measures, such as masks, plexiglass partitions, greater distance between assessor and client, etc., which have no empirical inquiry). Our research evidence base is young and imperfect, but it allows for some confidence in some measures administered in this way.

STATE OF THE RESEARCH

The research landscape focused on psychological tele-assessment has amassed a modest, though compelling, body of evidence for the reliability, validity, and utility of tele-assessment procedures in collecting psychological data. First, as a primary measurement tool in psychological assessment, the evidence that *clinical interviews* conducted through tele-assessment procedures are generally equivalent to traditional, in-person procedures is quite strong (Garb, 2007; Hyler, Gangure, & Batchelder, 2005; Luxton, Pruitt, & Osenbach, 2014; Schopp, Johnstone, & Merrell, 2000; Singh, Arya, & Peters, 2007). The Society for Personality Assessment's (COVID-19 Task Force to Support Personality Assessment, 2020) guidance on tele-assessment of personality and psychopathology notes that this is likely related to the fact that clinical interview data accuracy is heavily related to the quality of the therapeutic alliance and relationship, which has been shown to be quite strong in telehealth in general (Bouchard et al., 2000; Germain, Marchand, Bouchard, Guay, & Drouin, 2010; Morgan, Patrick, & Magaletta, 2008; Simpson, 2001). The equal quality of data elicited through tele-assessment and in-person procedures seems especially true for more structured clinical interviews (Grady et al., 2011; Hyler et al., 2005; Ruskin et al., 1998; Shore, Savin, Orton, Beals, & Manson, 2007), and has even been supported in forensic evaluations (Lexcen, Hawk, Herrick, & Blank, 2006; Manguno-Mire et al., 2007). There is similarly relatively strong evidence that *self-report questionnaire measures* are generally equivalent (Garb, 2007; Luxton et al., 2014), especially if specific steps are taken to ensure the integrity of the self-report tests (Corey & Ben-Porath, 2020).

As stated previously, a great deal of the literature on equivalence between psychological tele-assessment and traditional, in-person psychological assessment administration of measures is found on *neuropsychological tests*. Research has evaluated the equivalency of specific, performance-based neuropsychological tasks and built a significant evidence base for tele-neuropsychological testing. Most of this work has included older adult populations and a variety of tasks. These tasks include broader, multitask and multi-construct measures like the Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) and the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS; Randolph, 1998), both of which include verbal and nonverbal components. Also evaluated are select tests and subtests that evaluate single neuropsychological constructs, like the Boston Naming Test (Kaplan, Goodglass, & Weintraub, 1976), Digit Span, Matrix Reasoning, and Vocabulary subtests of the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 2008), and drawing tasks like clock drawing and the Beery Visual-Motor Integration, Fourth Edition (VMI-I; Beery, 2004). The literature has shown that in general these performance-based, tele-neuropsychological assessment techniques are effective at evaluating older adults, both with and without cognitive impairment, and discriminating impaired from non-impaired individuals equally as well as in-person neuropsychological procedures (Cullum, Weiner, Gehrmann, & Hynan, 2006; Galusha-Glasscock, Horton, Weiner, & Cullum, 2016; Grosch, Weiner, Hynan, Shore, & Cullum, 2015; Harrell, Wilkins, Connor, & Chodosh, 2014; Loh, Donaldson, Flicker, Majer, & Goldswain, 2007; Luxton et al., 2014; Temple, Drummond, Valiquette, & Jozsvai, 2010; Tukstra, Quinn-Padron, Johnson, Workinger, & Antoniotti, 2012; Wadsworth et al., 2018).

In a systematic review and meta-analysis of neuropsychological tele-assessment test administration with adults, Brearly and colleagues (2017) identified several themes, including overall findings that videoconferencing administration did not yield any significant change in test scores when compared to in-person assessment, with a very small effect size noted for administration procedure. They noted an average of 1/33rd of a standard deviation lower scores elicited by tele-assessment procedures than by in-person ones. They did note in their analyses that while verbally mediated, synchronous tests (those that require in-the-moment interaction and scoring) were unaffected by the different administration modalities, those tests with both verbal and visual stimuli did have significantly lower scores in tele-assessment conditions. However, the effect size of these differences, even though significant, was small, and scores were about 1/10th of a standard deviation lower in the tele-assessment condition. It should be noted that most equivalence research on neuropsychological tests has taken a task-based (rather than full measure-based) approach.

Rapid References 1.1 and 1.2 summarize the major studies that have attempted to determine equivalence between traditional in-person methods and tele-assessment methods of test administration (to date). In addition to study details,

DON'T FORGET

Study Ratings in Tables 1.1 and 1.2 used the following criteria:

a = peer reviewed
b = convincing sample size
c = strong research design and statistical analyses to support equivalence

The ratings are as follows (and the reason for a downgraded rating is listed using these letter codes to represent problematic areas in the specific study, listed in the table under "Rating Reason"):

1 = All three criteria (a, b, and c) met
2 = Two of three criteria met
3 = One of three criteria met

a study rating column is provided to rate the strength of each study. The studies are rated on a three-point scale, with 1 indicating the strongest study qualities (and thus the most convincing evidence) and 3 indicating less strength. To receive a study rating of "1," a study must (a) be peer reviewed, (b) possess a convincing sample size, and (c) provide a strong research design (e.g., random assignment to groups) and statistical results (e.g., not merely a correlation or interrater reliability) that allow equivalence to be examined. If any one of these three criteria were not met, the study received a rating of 2. If two of these three criteria were not met, the study received a rating of 3 (studies that did not meet any of these criteria were not included). The Rating Reason column provides the area deemed problematic (i.e., a, b, or c in the aforementioned description) that resulted in the Study Rating being downgraded.

While not nearly as much has been accomplished in areas other than neuropsychological tele-assessment, there has been some effort to evaluate equivalence between traditional and tele-assessment administration procedures for *cognitive/intellectual ability tests*. While some of the neuropsychological tele-assessment literature specifically looked at subcomponents of cognitive tests like the WAIS and similar tasks (Brearly et al., 2017; Temple, Drummond, Valiquette, & Jozsvai, 2010), only a few studies have been completed looking at equivalence between full cognitive ability and academic achievement measures in tele-assessment and traditional assessment contexts. Hodge and colleagues (2019) conducted a relatively small field study to determine agreement between scoring on traditional administration and tele-assessment administration of the Wechsler Intelligence Scale for Children, Fifth Edition (WISC-V; Wechsler, 2014) with students with learning disabilities. They found extremely high correlations for the different WISC-V primary index scores (ranging from .981 to .997), with the Full Scale IQ correlated at .991.

In larger studies, Wright (2018a, 2018b, 2020b) found similar evidence of (mostly) no significant impact of tele-assessment versus in-person methodology

Rapid Reference 1.1

Tele-Assessment Mode Equivalence Studies for Neuropsychological Tasks

Reference	Population	Study Rating	Rating Reason	*N*	Ages	Measures/Tasks	Software	Data	Results
Barcellos, Bellesis, Shen, Shao, & Chinn et al., 2018	Multiple sclerosis	1	n/a	270	18–69	California Verbal Learning Test-II	n/a (Telephonic)	Mean for 5 trials	Equivalent
Brearly et al., 2017	Meta-analysis of healthy, medical, psychiatric, substance use, dementia, and mild cognitive impairment (MCI)	1	n/a	497, 12 studies	Adult	Numerous, but multiple studies aggregated for: – Boston Naming Test – Semantic fluency – Clock Drawing – List learning (total) – MMSE – Phonemic fluency	Videoconference	Hedges' g (standardized mean difference)	– Equivalent overall – Of 12 accepted studies, one with significant mode effect, and two with Hedges' g ≥ .20 (criteria for small effect) – Boston Naming Test had significant mode effect – No task-specific analyses, age, or population-specific analyses yielded Hedges' g ≥ .20
Cullum et al., 2006	Neurocognitive (MCI, Alzheimer's)	2	b	33	51–84	– MMSE – Hopkins Verbal Learning Test – Clock Drawing Test – Digit Span – Category Fluency – Letter Fluency – Boston Naming Test (short form)	Videoconference	– Task-level means – Correlations between modes	All equivalent except Hopkins Verbal Learning Test retention percentage and Clock Drawing – Correlations ranged from .54–.88

(continued)

Reference	Population	Study Rating	Rating Reason	*N*	Ages	Measures/Tasks	Software	Data	Results
Cullum, Hynan, Grosch, Parikh, & Weiner, 2014	Mixed (MCI, Alzheimer's, healthy)	1	n/a	202	46–90	– MMSE – Hopkins Verbal Learning Test-Revised – Digit Span forward and backward – Boston Naming Test (short form) – Letter Fluency – Category Fluency – Clock Drawing	Videoconference	– Task-level means – Correlations between modes	– Means statistically equivalent except Clock Drawing Test and Hopkins Verbal Learning Test-R, but differences deemed clinically negligible (BNT = .2; HVLT = .9) – Correlations ranged from .55–.91 with mean of .74
Dekhtyar, Braun, Billot, Foo, & Kiran, 2020	Aphasia	2	b	20	26–75	Western Aphasia Battery–Revised (WAB-R)	Teleconferencing (consumer grade)	Composite means, interrater reliability, and intercorrelations across modes	Equivalent
Galusha-Glasscock et al., 2016	Mixed (MCI, Alzheimer's, healthy)	2	b	18	58–84	RBANS index scores: – Immediate Memory – Visuospatial/Constructional – Language – Attention – Delayed Memory – Total Scale	Videoconference	– Index score means – Correlations between modes	– Means statistically equivalent – Correlations ranged from .59–.90 with mean of .80; all significant
Grosch et al., 2015	Psychiatric outpatient	2	b	8	67–85	– MMSE – Clock Drawing Test – Digit Span	Videoconference	– Task means – Correlations between modes	– Means statistically equivalent – Correlations ranged from .42–.72

Hildebrand, Chow, Williams, Nelson, & Wass, 2004	Healthy	2	b	29	60 +	Rey Auditory Verbal Learning Test Controlled Word Association Test WAIS-III and WASI Vocabulary WAIS-III and WASI Matrix Reasoning Brief Test of Attention Clock Drawing	Videoconference	– Task means – Correlations between modes	Equivalent except Clock Drawing Test
Jacobsen, Sprenger, Andersson, & Krogstad, 2003	Healthy	2	b	32	Mean: 35	– Benton Visual Retention Test – Digit Span – Grooved Pegboard – Symbol Digit Motor Test – Seashore Rhythm Test – Visual Object and Space Perception Battery Silhouettes – WAIS Vocabulary and Digit Span (Norwegian) – Wechsler Memory Scale (WMS) Logical Memory	Videophone	– Task means – Correlations between modes	– Equivalent except WMS Logical Memory I and Seashore Rhythm Test (auditory attention) – Correlations ranged from .37–.86 with median of .74

(continued)

Reference	Population	Study Rating	Rating Reason	*N*	Ages	Measures/Tasks	Software	Data	Results
Stain et al., 2011	Psychosis	2	b	11	14–27	– Wechsler Test of Adult Reading – WMS Logical Memory – WAIS-III Digit Span – Controlled Oral Word Association Test (verbal fluency)	Videoconferencing	– Task means – Correlations between modes	– Equivalent except Wechsler Test of Adult Reading – Correlations ranged from .59–.96 with average of .84
Sutherland 2017	Language impairment and SLD	3	b, c	23	8–12	CELF-4 subtests and composites: – Concepts and Following Directions – Word Structure – Recalling Sentences – Formulated Sentences – Word Classes – Core Language Score – Receptive – Expressive	Telehealth application National Information Communications Technology Australia (NICTA)	– Composite interrater correlations between modes	– Correlations ranged from .96–1.0
Turkstra et al., 2012	Traumatic brain injury (moderate to severe)	2	b	20	21–69	– Mediated Discourse Elicitation Protocol – AphasiaBank discourse tasks	PCbased Polycom encrypted videoconferencing or iChat and iSight on Mac Powerbooks	– Task means – Correlations between modes	– Equivalent – Overall correlation was .92

Vahia et al., 2015	Older rural Latino adults with suspected cognitive impairment	2	b	22	65 +	– MMSE – Hopkins Verbal Learning Test – EIWA-III Digit Span – Letter Fluency – Category Fluency – Clock Drawing – Brief Visuospatial Memory Test – PontonSatz Spanish Naming Test – Standardized measure of overall cognitive functioning	Videoconferencing Remotely controlled Pan Tilt and Zoom cameras	– Task means – Overall cognitive functioning mean	– Equivalent
Vestal, Smith-Olinde, Hicks, Hutton, & Hart, 2006	Alzheimer's	2	b	10	68–78	– Picture Description (Boston Diagnostic Aphasia Examination) – Boston Naming Test – Token Test (Multilingual Aphasia Examination) – Aural Comprehension of Words and Phrases (Benton) – Controlled Oral Word Association Test (Benton)	Polycom videoconferencing	Task means	Equivalent

(continued)

Reference	Population	Study Rating	Rating Reason	*N*	Ages	Measures/Tasks	Software	Data	Results
Wadsworth et al., 2018	Cognitively impaired and nonclinical	1	n/a	197	Means: 66 healthy 73 impaired	– MMSE – Hopkins Verbal Learning Test – Letter Fluency – Category Fluency – Boston Naming Test – Digit Span forward and backward – Clock Drawing Test	Polycom videoconferencing	Task means	Equivalent except Category Fluency (small effect not deemed clinically meaningful)
Waite, Theodoros, Russell, & Cahill, 2010	Language impairment or difficulties	3	b, c	25	5–9	CELF-4: Concepts and Following Directions Word Structure Recalling Sentences Formulated Sentences	Videoconferencing Scanned stimuli Stimulus video recordings	Intra- and interrater reliability	Equivalent

WAB-R subtests: Conversational Questions, Picture Description, Yes/No Questions, Auditory Word Recognition, Sequential Commands, Object Naming, Word Fluency, Sentence Completion, Responsive Speech, Reading Comprehension of Sentences, Reading Commands, Written Word–Object Choice Matching, Written Word-Picture Choice Matching, Picture-Written Word Choice Matching, Spoken Word-Written Word Choice Matching, Spelled Word Recognition, Spelling, Writing Upon Request, Writing Output, Writing to Dictation, Writing Dictated Words, Alphabet and Numbers, Dictated Letters and Numbers, Copying a Sentence, Drawing, Block Design, Calculation, Raven's Coloured Progressive Matrices, Writing Irregular Words to Dictation, Writing Nonwords to Dictation, Reading Irregular Words, Reading Nonwords. Composites: Aphasia Quotient, Language Quotient, Cortical Quotient.

RBANS subtests: List Learning, Story Memory, Figure Copy, Line Orientation, Picture Naming, Semantic Fluency, Digit Span, Coding, List Recall, List Recognition, Story Recall, and Figure Recall.

in the administration of three other cognitive tests, the Woodcock–Johnson IV Tests of Cognitive Ability (WJ-IV-Cog; Schrank, McGrew, & Mather, 2014), the Reynolds Intellectual Assessment Scales, Second Edition (RIAS-2; Reynolds & Kamphaus, 2015), and the WISC-V. In each similarly designed study, clients (children/students) were recruited, matched on demographic characteristics (at least age and gender), then randomly assigned to either the control (traditional, in-person administration) or experimental (tele-assessment) condition. In the WJ-IV-Cog study, clients were also administered the Cognitive Abilities Test (CogAT; Lohman & Hagen, 2001), all in the same format, as a randomization check. For the WISC-V study, clients were administered the Kaufman Brief Intelligence Test, Second Edition (KBIT-2; Kaufman & Kaufman, 2004) in traditional format for the same purpose. The WJ-IV and the RIAS-2 studies reported both significance (*p*) and effect sizes, and both found negligible to small effects (<.20), all non-significant, for administration method. The WISC-V study utilized a different approach to equivalence testing that is well known across many industries (see Chapter 4), reporting confidence intervals for the subtest and composite score differences. This study also supported equivalence on all scores, except for one secondary subtest that exceeded the predetermined criteria for a small effect (<.30). While some have identified that, despite revealing no statistical differences, the samples did exhibit actual score differences (especially on the RIAS-2; Farmer et al., 2020), this ignores the purpose of evaluating the statistical equivalence of the methods. Of course, there will be sample variations in actual scores obtained, but the statistical equivalence between the two methods takes into account error in measurement and non-identical samples (see Chapter 4 for further discussion on equivalency).

To date, only one equivalency study of tele-assessment and traditional assessment of *academic achievement tests* was identified. Wright's (2018b) study of tele-assessment procedures for the Woodcock–Johnson IV Tests of Achievement (WJ-IV-Ach; Schrank et al., 2014) similarly found no statistical significance or notable effect sizes (all were well below .20) for assessment procedure. Like the study of the WJ-IV-Cog, the WJ-IV-Ach study included a relatively large sample size (240 students in total, split across conditions), a design to ensure equivalence between groups (case control matching on age and gender), and randomization (matched pairs were randomly assigned to condition, and a randomization check utilizing the CogAT measure was utilized). This study begins a foundation for establishing that academic achievement tests can be given through tele-assessment procedures, maintaining their reliability, validity, and utility.

It must be noted that the overwhelming majority of the research on equivalence between tele-assessment and traditional, in-person procedures for specific test administration has been conducted in highly controlled environments and most often with the use of a proctor or helper of some sort on the client's side of the assessment. That is, very few of the studies anticipated the need for individuals to be assessed in their homes without the ability to dispatch some sort of (even non-psychologist) helper. Additionally, there are only studies on a few very specific measures.

However, these studies provide a foundation to support the conclusion that results derived from tasks with previously studied input and output requirements—and thus the psychological, cognitive, behavioral, and interpersonal processes that the tasks engage—are not significantly altered when administered in a tele-assessment format. For example, an auditory digit span task (which has shown to be equivalent across multiple studies and sample types) requires an auditory verbal stimulus and an auditory verbal response. Other measures that require these inputs and outputs (and especially those evaluating the construct of working memory) have indirect evidence of equivalence. Similarly, a writing task exhibiting equivalence on the WJ-IV-Ach should indirectly support a similar writing task on another test, as the inputs, construct-relevant processes evaluated, and outputs are extremely similar. Rapid Reference 1.2 summarizes the existing tele-assessment mode equivalence studies that have been conducted on cognitive, achievement, and neuropsychological tests to date.

DON'T FORGET

Existing equivalence studies support the conclusion that results derived from tasks with previously studied input and output requirements are not significantly altered when administered in a tele-assessment format.

The two areas of general psychological assessment practice that (at the time of writing this book) have extremely limited guidance and support, either in the research or from professional organizations, are forensic assessment and developmental—specifically autism-spectrum disorder (ASD)—assessment. Specifically for *forensic assessment*, Drogin (2020) highlights that courts have not yet truly engaged in the discussion of tele-assessment and its use in forensic settings, but they are likely going to have to quite soon. He also notes, though, that it is very likely that attorneys on both sides will soon become extremely good at both defending and criticizing tele-assessment work. A great deal of forensic assessment does not include formal performance-based testing, and structured and semi-structured forensic interviews have some support (Lexcen et al., 2006; Manguno-Mire et al., 2007), but the field will need to reconcile this specific difficulty relatively quickly and soon.

Rapid Reference 1.2

Tele-Assessment Mode Equivalence Studies for Cognitive and Academic Measures

Reference	Population	Study Rating	Rating Reason	*N*	Ages	Measures/Tasks	Software	Data	Results
Hodge et al., 2019	Specific learning disorder (SLD)	3	b, c	33	8–12	WISCV primary index scores and Full Scale IQ	Coviu	– Composite inter-rater correlations between modes	– Correlations ranged from .98–.99
Temple et al., 2010	Intellectual disability	2	b	19	23–63	– Beery–Buktenica Developmental Test of VMI WASI composites: – VIQ – PIQ – FSIQ	Polycom encrypted videoconferencing	– Composite means – VMI mean – Correlations between modes	– Mean differences ranged from – 0.6–2.1 standard score points – VIQ significantly different at .05 level – Correlations ranged from .92–.98
Wright, 2018a	General	2	a	104	3–19	RIAS2 (8 subtests and all composites)	PresenceLearning	Subtest and composite means	Equivalent except Speeded Naming Task (equivalent at ages 7–19 only) and Speeded Processing Index

(continued)

Reference	Population	Study Rating	Rating Reason	*N*	Ages	Measures/Tasks	Software	Data	Results
Wright, 2018b	General	1	n/a	240	5–16	WJ IV Cog tests and composites (10 subtests) WJ IV Ach tests and composites (11 subtests)	PresenceLearning	Test and composite means	Equivalent
Wright, 2020b	General	1	n/a	256	6–16	WISC-V (10 primary and 6 secondary subtests) WISC-V primary index scores and FSIQ	PresenceLearning	– Subtest and composite means – Subtest and composite intercorrelations within and between modes	Equivalent except Letter–Number Sequencing

RIAS-2 subtests: Guess What, Odd-Item Out, Verbal Reasoning, What's Missing, Verbal Memory, Nonverbal Memory, Speeded Naming, and Speeded Picture Search. Composites: Verbal Intelligence Index, Nonverbal Intelligence Index, Composite Intelligence Index, Composite Memory Index, Speeded Processing Index.

WISC-V primary subtests: Similarities, Vocabulary, Block Design, Visual Puzzles, Matrix Reasoning, Figure Weights, Digit Span, Picture Span, Coding, Symbol Search. Secondary subtests: Information, Comprehension, Picture Concepts, Arithmetic, Letter–Number Sequencing, Cancellation. Primary index scores: Verbal Comprehension Index, Visual Spatial Index, Fluid Reasoning Index, Working Memory Index, Processing Speed Index.

WJ-IV-Cog tests: Oral Vocabulary, Number Series, Verbal Attention, Letter–Pattern Matching, Phonological Processing, Story Recall, Visualization, General Information, Concept Formation, Numbers Reversed. Composites: General Intellectual Ability, *Gf-Gc*, Comp-Knowledge, Fluid Reasoning, Short-term Working Memory. Cognitive Efficiency.

WJ-IV-Ach tests: Letter-Word Identification, Applied Problems, Spelling, Passage Comprehension, Calculation, Writing Samples, Word Attack, Oral Reading, Sentence Reading Fluency, Math Facts Fluency, Sentence Writing Fluency. Composites: Broad Reading, Broad Mathematics, Broad Writing.

When it comes to *ASD assessments*, no professional organization has yet (as of the time of writing this book) made a statement or offered guidance on developmental tele-assessment. This is especially notable as the Autism Diagnostic Observation Schedule, Second Edition (ADOS-2; Lord et al., 2012) is widely accepted as a gold-standard assessment of ASD, and it includes so many physical toys and manipulatives and so much specific interaction that it is nearly impossible to complete via tele-assessment procedures (with the possible exception of some of the interview prompts in Module 4, for older clients). While the ADOS-2 is the most widely used performance-based measure to evaluate ASD, there are certainly many other tests, including interview measures (that primarily focus on parent/caregiver report) and holistic clinician rating measures (that are meant to account for amassed evidence from different sources). Chapter 8 of this book discusses some of these measures in detail.

CURRENT PROFESSIONAL GUIDANCE

As of the writing of this book, only a handful of professional organizations have provided specific guidance documents related to conducting psychological assessment in a telehealth modality. The guidance offered by these professional organizations forms the basis of the recommendations used throughout this book. However, much of it is vague and aspirational. Presently, several organizations have emerged with specific guidance related to the practice of tele-assessment (Table 1.3). The American Psychological Association (APA)—in addition to their general guidelines for the practice of telepsychology (Joint Task Force for the Development of Telepsychology Guidelines for Psychologists, 2013)—has published guidance for tele-assessment in general (Wright, Mihura, Pade, & McCord, 2020), as well as specific guidance for tele-assessment with children (Banks & Butcher, 2020), pain evaluations (Brown & Bruns, 2020), and presurgical evaluations (Block, Bradford, Butt, & Marek, 2020). The Inter Organizational Practice Committee (IOPC, 2020a, 2020b) has produced two separate documents related to the practice of neuropsychological tele-assessment. The National Association of School Psychologists (NASP) developed and published a general guidance document in 2017 and offered a specific update related to the COVID-19 crisis (NASP, 2020). The Society for Personality Assessment (SPA) published guidance on tele-assessment of personality and psychopathology (COVID-19 Task Force to Support Personality Assessment, 2020). Although there have been no professional guidance documents related to forensic work, the California Commission on Peace Officer Standards and Training (2020) disseminated guidance on pre-employment tele-assessment for police officers (see Rapid Reference 1.3).

Rapid Reference 1.3

Professional Guidance and Recommendation Documents for Psychological Tele-Assessment

APA General Guidance on Tele-Assessment: https://www.apaservices.org/practice/reimbursement/health-codes/testing/tele-assessment-covid-19

IOPC Guidance for Teleneuropsychology: https://static1.squarespace.com/static/50a3e393e4b07025e1a4f0d0/t/5ed7d6c58ec40f3dce143b40/1591203525610/IOPC+Models+of+Care+During+COVID-19+Pandemic.pdf

NASP Guidance for Delivery of School Psychological Telehealth Services: http://www.nasponline.org/assets/documents/Guidance_Telehealth_Virtual_Service_%20Delivery_Final%20(2).pdf

NASP Guidance on Virtual Service Delivery: 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

SPA Guidance on Tele-Assessment of Personality and Psychopathology: https://resources.personality.org/www.personality.org/General/pdf/SPA_Personality_Tele-Assessment-Guidance_6.10.20.pdf

APA Guidance on Tele-Assessment of Children: https://www.apaservices.org/practice/legal/technology/telehealth-testing-children-covid-1erican

APA Guidance on Tele-Assessment for Presurgical Evaluations: https://www.apaservices.org/practice/news/presurgical-psychological-evaluations-covid-19?_ga=2.116775942.169813268.1591026574-1626901323.1573678255

APA Guidance on Tele-Assessment for Chronic Pain Evaluations: https://www.apaservices.org/practice/news/chronic-pain-covid-19

California Commission on Peace Officer Standards and Training Guidance on Tele-Assessment with POST Selection Standards: https://post.ca.gov/Portals/0/post_docs/bulletin/2020-18.pdf

Most of these professional guidance documents offer overlapping advice, though some are much more specific than others. Generally, they acknowledge the limitations of the current literature and evidence base for tele-assessment mode equivalence with traditional in-person administration. They also tend to acknowledge that psychologists have many resources available to them to conduct assessments, including multiple methods, early support for administration mode equivalence, and the ability to adjust interpretations appropriately, utilizing clinical and professional judgment, training, and experience.

While they do not outright say this, by acknowledging that tele-assessment can be a viable procedure for collecting psychological data, they imply that imperfect data can still be good, useful data. While bad data is worse than no data at all, data collected through tele-assessment procedures can be imperfect but good, usable data.

DON'T FORGET

Imperfect data can still be good, useful data. Data collected through tele-assessment procedures can be imperfect but good, usable data.

MEASURES WITH LONGSTANDING TELE-ASSESSMENT ADMINISTRATION PROCEDURES

Many psychological assessment measures have a significant history of tele-assessment administration. The most obvious of these are collateral report questionnaires and surveys. While many of these began as physical, pencil-and-paper questionnaires often sent home with clients to give to knowledgeable others to fill out, more recently, most of these have adapted to a remote, online administration procedure. Surveys for parents and teachers, for example, to fill out are widely administered in this fashion. A link is sent out to the respondent, who fills out the survey online wherever they are and whenever they want, and the report is generated for the clinician. There are even some self-report inventories that have been developed and utilized in this way, without very much concern for test security or integrity. It is important to note that these measures constitute a part of an assessment that carries with it *absolutely no change at all* in the "new" landscape of tele-assessment. The conversation about psychological tele-assessment should always acknowledge that at least some of the "traditional" work we have been doing has included methodologies that can be accomplished entirely unaltered.

MEASURES DEVELOPED FOR ONLINE ADMINISTRATION

While much of the focus in psychological tele-assessment is on how to adapt traditionally developed in-person measures to an online, virtual procedure, there have been some (and will likely continue to be more) measures that were specifically developed, standardized, and normed for online use. Aside from questionnaires and surveys, several cognitive and neuropsychological measures have been developed for online use and may prove useful, though each has some significant drawbacks, mostly in terms of the state of the supportive empirical evidence. However, each of these represents some promising moves in the field that are likely to improve our tools in the future.

One promising general measure of cognitive ability for ages 6 through adult is the Mezure instrument (mezureschools.com). Built around measures of crystallized (Gc) and fluid (Gf) intelligence, it includes online tasks for these constructs as well as processing speed, memory with distractions, social perception, and (for adults) stress tolerance. The development of the test is impressive, as are the data security features and graphical and auditory stimuli, and many of the tasks do seem to tap what they are intended to measure. The clinical manual (Mezure, n.d.) provides solid test development information, including reliability and some validity information, such as an exploratory factor analysis. The standardization sample is adequately large at 4184 individuals, though they break down the numbers by age band for children, but not adults (and it is unclear if or how they standardized for older adults). The biggest weakness for the Mezure as it currently stands is in its presentation of validity information. The criterion-related validity (only concurrent) is limited to a correlation between the overall score of the Mezure (based on the seven core subtests) and the WISC-III (Wechsler, 1991) Full Scale IQ, Verbal IQ, and Performance IQ. The three major flaws with this are that: 1) the WISC-III is obviously quite outdated (and there is no indication of when this study was performed); 2) this study only addresses validity for children aged 6 to 16, with no evidence of validity for use with adults, which the Mezure claims to work for as well; and 3) the relationships between subtests of the Mezure and similar constructs on other tests are not reported, even though presumably the data exist from the WISC-III study. Similarly, the supplemental subtests (processing speed, social perception, distractibility) lack reported validity evidence. Further, internal validity is only asserted through the use of subtest correlations, rather than a confirmatory factor analysis. Until these data are provided, it is recommended that the Mezure be used with extreme caution.

Another measure entirely developed and standardized online to evaluate general cognitive ability for individuals age 7 through adult is the Cognifit measure (cognifit.com). This measure reportedly evaluates individuals on focus, distractibility, processing speed, spatial perception, inhibition, and a host of other cognitive abilities, all organized into perception, coordination, attention, memory, and reasoning. Most of the subtests are generally quite technologically elegant and seem to measure what they purport to measure. There are two significant drawbacks to this program, though. First, the report generated is extremely difficult for psychologists who understand psychometrics to interpret. Specifically, the program has developed an odd scoring system (it is unclear if this is even a standard score) with a maximum of 800, making it extremely difficult to interpret. Additionally, the algorithms used to derive scores on indices are somewhat obtuse and not well explained in any materials (individual subtests contribute partially to multiple different index scores). The second and more troubling problem with the measure is its description of psychometrics (CogniFit, 2020). The only psychometric information provided in the clinical manual is measures of reliability (which it calls validity), both internal consistency and test-retest. While the reliability findings are adequate, they are reported for the tasks (not the derived index scores at all), and no validity information has been provided. There is no information on the internal structure of the test itself, nor is there evidence of concurrent or predictive criterion-related validity. As such, this measure is not yet ready to be used by professionals (and indeed seems to be marketed more to lay people and medical professionals).

Finally, the CNS Vital Signs measure (cnsvs.com) offers another online-developed and standardized instrument of cognitive functioning for individuals aged 7 to 90. The measure reportedly evaluates memory, fine motor functioning, attention and concentration, processing speed, and mental flexibility. The measure does not purport to evaluate general cognitive functioning or intelligence, but rather employs some commonly used neuropsychological techniques to evaluate discrete neuropsychological and cognitive abilities. While research (summarized in Gualtieri & Johnson, 2006) has established adequate reliability, the normative sample is quite small (25 children under 10 years old) and not very racially or ethnically diverse. The validity data are generally adequate but based again on quite small sample sizes, especially related to discrimination ability with certain neurocognitive disorders (e.g., 52 participants in a dementia discrimination study). In all, there are fewer concerns with a measure like the CNS Vital Signs than the other measures, as it does not purport to measure broader general cognitive abilities, but rather discrete skills that are reasonably well established in neuropsychological and tele-neuropsychological assessment, though caution should be taken especially with children.

CONCLUSION

The current landscape of psychological tele-assessment is either promising but young at the moment or young but promising at the moment, depending on perspective. While imperfect, in general the professional organizations that organize health service psychologists (school, clinical, counseling, and related fields) have encouraged careful and considered practice when engaging in tele-assessment. And the research seems to support this—some tests have direct evidence supporting their equivalence and thus use for tele-assessment, and many more can use indirect evidence from the necessary inputs (stimuli) and outputs (client responses) that have been evaluated to support their use as well. Further, while currently none of the measures is ideal for use yet, several cognitive measures have been developed and standardized entirely online, and this sets the stage for an exciting future of psychological tele-assessment and further development of fully online measures.