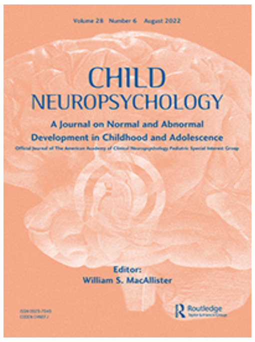

Child Neuropsychology

A Journal on Normal and Abnormal Development in Childhood and Adolescence

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ncny20

Tele-assessment of cognitive functions in children: a systematic review

Costanza Ruffini, Christian Tarchi, Monica Morini, Gabriella Giuliano & Chiara Pecini

To cite this article: Costanza Ruffini, Christian Tarchi, Monica Morini, Gabriella Giuliano & Chiara Pecini (2022) Tele-assessment of cognitive functions in children: a systematic review, Child Neuropsychology, 28:6, 709-745, DOI: 10.1080/09297049.2021.2005011

To link to this article: https://doi.org/10.1080/09297049.2021.2005011

Published online: 02 Dec 2021.

CHILD NEUROPSYCHOLOGY
2022, VOL. 28, NO. 6, 709–745
https://doi.org/10.1080/09297049.2021.2005011

REVIEW ARTICLE

CE Tele-assessment of cognitive functions in children: a systematic review

Costanza Ruffini, Christian Tarchi, Monica Morini, Gabriella Giuliano and Chiara Pecini

Department of Education, Languages, Intercultures, Literatures and Psychology (FORLIPSI), University of Florence, Firenze, Italy

ABSTRACT

Cognitive Tele-Assessment approach (CTA) has been widely used in adults for clinical, research, and screening purposes. In the last decades, it has been considered a useful tool for evaluating child development in both clinical and educational settings and new instruments for CTA in children have been developed. In comparison to In Person Assessment (IPA), CTA can have several advantages, such as increasing accessibility, cutting waiting lists, reducing time and travel costs, and assisting with infection control by minimizing face-to-face contact in times of pandemic. Nevertheless, several issues related to the feasibility and reliability of using CTA to evaluate cognitive development are still open. The present systematic review has a twofold aim: 1. to describe the cognitive functions that are most frequently measured by CTA in children, the procedures used, and the characteristics of the samples investigated; 2. to investigate the agreement between CTA and IPA scores in children.

In the present systematic review, 23 studies using CTA in children, with typical or atypical development, have been selected and analyzed. Results support the similarities in performance scores between IPA and CTA and good compliance by children and their families in participating in CTA. Nonetheless, most studies suggest that several methodological precautions must be taken to manage technical and procedural characteristics that may represent challenges for CTA of children. Suggestions for a correct use of CTA, factors affecting the validity of the results and directions for future research are discussed.

ARTICLE HISTORY
Received 23 June 2021
Accepted 8 November 2021

KEYWORDS
Remote assessment; tele-assessment; child; cognition; development

Introduction

In the last few decades, the scientific panorama has highlighted the utility of Cognitive Tele-Assessment (CTA) that is a type of assessment of cognitive functioning in which examiners and participants interact through telecommunications technologies rather than being in the same place (Hodge et al., 2019a; Krach et al., 2020). Further interest in CTA has recently arisen due to the physical distances imposed by Covid-19 emergency that required, for a protracted period, the interruption or substitution of traditional face-

CONTACT Costanza Ruffini costanza.ruffini@unifi.it Department of Education, Languages, Intercultures, Literatures and Psychology (FORLIPSI), University of Florence, Via Di San Salvi 12, Firenze 50135, Italy

to-face interactions in all research and clinical fields, both in adults and children (Farmer et al., 2020). Therefore, it is of paramount importance to describe the results of studies using tele-assessment and provide recommendations and guidelines for future clinical and research procedures. This need is even more pronounced for children because of their higher variability in engagement, attention, and compliance in comparison to adults (Hodge et al., 2019a; Manning et al., 2020; Raman et al., 2019; Salinas et al., 2020).

This article aimed at reviewing existing studies of CTA in children in order to:i) report the cognitive functions that are most frequently measured by CTA in children, the procedures used and the characteristics of the samples investigated; and ii) investigate the agreement in children's scores between CTA In Person Assessment (IPA).

Information and communication technologies in e-Health

Since the 1960s, e-health, that is any use of information and communication technologies (ICT) in the health system, has been progressively introduced in clinical and educational practices (Hodge et al., 2019a; Raman et al., 2019). Moreover, specific academic associations, such as the American Telemedicine Association, have been founded. Indeed, the use of ICT to provide health care remotely (Jessiman, 2002; Nagel & Penner, 2016) addresses accessibility issues due to distance and/or different family and special needs (Ciccia et al., 2011; O'Brian et al., 2014; Pearl et al., 2014). The use of tele-health tools has notably increased both for treatment (i.e., tele-rehabilitation, tele-intervention) and assessment (i.e., tele-assessment, online assessment, videoconferencing assessment) of cognitive functioning. This refers to the mental processes involved in receiving, selecting, storing, processing, and retrieving information from the environment, and include intelligence, language, learning, memory, attention, executive functions, perception, visuo-spatial abilities, motor-praxis functions, and academic skills (Lezak, 1995; Vallar & Papagno, 2007).

ICTs are considered feasible and reliable devices for administering cognitive assessment as, in comparison to traditional paper and pencil-standardized tests, they may provide multi-sensorial and motivating environments, thus increasing compliance and participation, especially in children (Luciana & Nelson, 2002). Moreover, nowadays, given that technologies are commonly used in everyday context, the cognitive assessment by ICT can be considered an ecologically valid tool to evaluate human mental functioning in specific circumstances. Finally, ICT opens to the possibility to obtain digital data on children with special educational needs to be used for clinical and diagnostic purposes. The *Automated Working Memory Assessment* (AWMA; Alloway, 2007) and the *Cambridge Neuropsychological Test Automated Battery* (CANTAB; Green et al., 2019; Luciana & Nelson, 2002) are two well-known examples of batteries using ICT to evaluate cognitive functioning in children, with typical and atypical development. AWMA assesses working memory and CANTAB assesses several cognitive processes. Both batteries require minimal training because the administration and scoring are fully automated and test scores are calculated directly by the software. Stemming from these two examples of ICT batteries to assess cognitive development, Luciana and Nelson (2002) have identified a few challenges in the use of ICT: a tendency to a compulsory use of the touchscreen in young children, a variability in the computer characteristics, and an intrusion of confounding variables like the lack of motivation due to the absence of a direct human relationship.

Nevertheless, the use of CTA in children has several benefits mainly related to the possibility of overcoming geographical, cultural, socio-economical and psychological barriers. For instance, in the USA, 15% of children do not receive a timely diagnosis because of both distance and cultural factors (Cortiella & Horowitz, 2014; Wright, 2020); in such circumstances, connecting remotely may be the only possible modality (Eriks-Brophy et al., 2008). CTA may also be useful when there is a shortage of experts in a range of locations or when the access to experts is restricted, as it is currently happening due to the pandemic (Corona et al., 2020; Sherwood & MacDonald, 2020; Wright, 2020). Moreover, CTA can meet the needs of children with medical, psychiatric, and physical conditions that make it challenging to travel or be away from their own environment for a long time (Harder et al., 2020; Ragbeer et al., 2016). Approximately 30 million people in the world fall in this category as they are affected by rare diseases and a huge number of other medical conditions. Finally, CTA offers ample opportunities to save time and money, as families may avoid long and expensive trips or taking time off work as well as clinicians may save scoring time, as the online tele-assessment platforms can automatically score child's responses and tests' output, thus, increasing the number of patients adequately assessed (Hodge et al., 2019a; Jessiman, 2002; Ragbeer et al., 2016). Another potential benefit is early diagnosis and early intervention (Corona et al., 2020; Eriks-Brophy et al., 2008; Juárez et al., 2018).

Over several advantages, there is still a public resistance to the use of CTA in children as it is assumed to have a research validity, but not a clinical one (May & Erikson, 2014). While the setting and the procedural rules for IPA are well defined and CTA on adults has been widely studied (Galusha-Glasscock et al., 2016; Wadsworth et al., 2018), there is a call for the development and refinement of guidelines for use of CTA in children, regarding the technology and personnel requirements, parameters of assessments and cognitive functions that can be assessed. For instance, the lack of a direct relationship between the examiner and the child, a fundamental characteristic of both the educational and clinical setting, may have several consequences for the feasibility and acceptability of CTA. Privacy and data protection too are fundamental issues to be considered during a CTA (American Psychological Association Services (APA Services), 2020b). A further variable that must be considered before choosing CTA is the presence of inequalities in the population in terms of the accessibility to technologies (Harder et al., 2020). Moreover, not all cognitive tests are suitable for tele-assessment as some of them may not be amenable to being presented remotely. Finally, the complexity of the web and the digital parameters of the CTA may impose detailed planning phases as well as along training for the examiners and the participants (Harder et al., 2020; Raman et al., 2019; Sutherland et al., 2017). Because of the brief history of CTA, these and other issues have not yet been systematically addressed and, most importantly, it is unclear whether norms and standardizations make CTA comparable to IPAin terms of reliability.

Rationale and objective of the review

To date, there are no reviews of the studies using tele-assessment to measure a broad range of cognitive functions in children. The two existing reviews in the field of CTA in children focused on the tele-assessment of speech and language (Edwards et al., 2012; Taylor et al., 2014) while the other cognitive functions were not considered. As CTA could be used in screening contexts, in clinical settings and in longitudinal research, a systematic review describing the main cognitive functions and identifying for whom and under what circumstances CTA may be appropriate, will contribute to defining good practices, strengths and weaknesses, as well as open issues for future research. The above are needed in order to avoid CTA becoming an extemporaneous practice in response to emergencies and pandemics.

Specifically, the following two research aims were pursued:

(1) to describe the cognitive functions most frequently measured by CTA in children and the variability in terms of procedures and sample characteristics;
(2) to investigate the validity and reliability of CTA in children in comparison to the in person assessment.

Methodology

Search strategy

The definition of the keywords was conducted between November and December 2020 and the comprehensive search on PubMed, PsycInfo, and Scopus databases and through the snowball technique was run on 5th of January 2021. The literature research was repeated on 7th of April 2021 to control for articles published since when the search was first performed, but no article meeting all the inclusion criteria was identified. A third literature search was conducted on 28th of October 2021 and three articles were included in the review as they met all the inclusion criteria. Keywords belonging to the following three clusters were used: children, tele-assessment and cognitive functions. The complete search string follows:

("pediatric*" or "child*" or "young children" or "school-aged children" or "youth") AND ("tele-assessment" or "telemedicine-based assessment" or "teleneuropsychology" or "videoconferencing" or "in-person-based assessment" or "remote assessment" or "telehealth" or "via telehealth" or "telepractice" or "telepsychology") AND ("cognitive" or "intelligence" or "intellectual abilit*" or "literacy" or "math" or "language" or "speech" or "memory" or "learning" or "attention" or "perception" or "visuo-spatial" or "motor" or "executive function*" or "neuropsycholog*")

The procedure details used for the different databases are reported in Appendix A.

Inclusion criteria below

(1) Use of tele-assessment, that is online remote evaluation by tele-communication tools;

(2) Tele-assessment including at least one cognitive measure (intelligence, literacy, math, language and speech, memory, learning, attention, perception, visuo-spatial abilities, motor skills, executive functions);
(3) Primary research articles;
(4) Papers written in English language;
(5) Samples including children aged between 18 months and 18 years.

Exclusion criteria below

(1) Not clear reference to cognitive tele-assessment;
(2) Reviews, books, guidelines, conference abstracts, pilot studies or commentaries;
(3) Articles written in other languages than English;
(4) Samples composed by only adults;
(5) Exclusive use of indirect measures of cognitive functions (i.g. questionnaires, inventories);
(6) Intervention studies without tele-assessment of cognitive functioning;
(7) Absence of a comparison with in person assessment procedures.

Study selection process

The research process is represented in Figure 1. It was carried out according to the PRISMA statements (Moher et al., 2009). Initially, the search gave 745 studies (PubMed n = 135, PsycInfo n = 154, Scopus n = 456) and one study was identified by reviewing the reference lists of the included articles. After removing duplicates, 535 studies remained, and they were screened for title and abstract by three independent evaluators (C.R., M. M., G.G.). On the basis of the title and/or abstract, 479 papers were excluded as they met at least one exclusion criteria: the agreement rate in this phase was 98%. The same evaluators independently read 56 full-text articles and in case of discrepancies, they discussed the reasons why an article could be included or excluded reaching a 100% agreement. Three articles were finally added to the review, as results of the third literature search. Twenty-three papers were selected.

Quality assessment

The quality assessment of the selected studies was conducted by applying the checklist for quantitative studies developed by Kmet et al. (2004). In order to control for risk of bias, the scale investigates 14 methodological characteristics: research question, study design, method, randomization of conditions, blinding of investigators and subjects, outcome measures, sample size, analytic methods, estimate of variance, control of confounding, detailed results and conclusion (see Appendix B). Each item has a 3-point answer: 2 points if the criterion is fully satisfied, 1 point if the criterion is partially satisfied, 0 points if the criterion is not satisfied. Some criteria may not be applicable to a specific study. The quality score is calculated for each single article by adding the scores of the items as obtained by each article and dividing them by the possible total score of the items actually

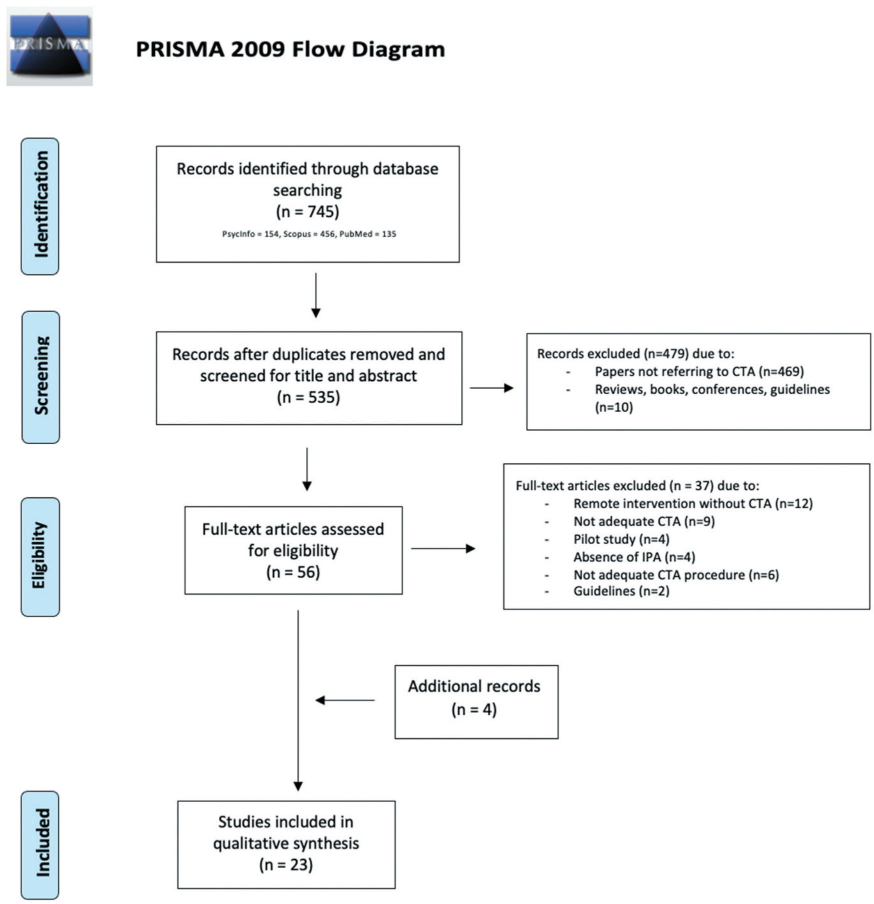

Figure 1. PRISMA flow diagram.

completed (excluding not applicable items). Three evaluators (C.R., C.P., G.G.) preliminary discussed the possible interpretations of each item and two of them (C.R., G.G.) ranked each study. A high level of inter-rater agreement (93.17%, 300 out of 322 answers) was obtained. The two evaluators discussed the discrepant items, finding an agreement on the score to be assigned. As reported in Appendix B, 13 studies got a high-quality score (>.80; Guiberson et al., 2015; Hamner et al., 2021; Harder et al., 2020; Hodge et al., 2019a, *Hodge, et al., 2019b; Kronenberger et al., 2021; Manning et al., 2020; Petrill et al., 2002; Sutherland et al., 2017; Waite et al., 2010a, 2010b, 2012; Wright, 2020), seven studies a medium-quality score (.80< rate <.60; Eriks-Brophy et al., 2008; Ragbeer et al., 2016; Raman et al., 2019; Salinas et al., 2020; Stain et al., 2011; Sutherland et al., 2019; Waite et al., 2006), and three a low-quality score (<.60; Ciccia et al., 2011; Jessiman, 2002; Werfel et al., 2021).

Results

The methodological characteristics and results of each study are reported in Table 1.

Table 1. Studies characteristics and results.

	Article	Sample (n)	Mean age (sd/ age range)	% male	Experimental conditions	Cognitive function	Remote administered outcome measure	Agreement CTA and in person assessment/scoring
1	Eriks-Brophy et al. (2008)	suspected of speech and language difficulties (7)	nr (4.3–12.9) yrs	nr	CTA vs in person scoring	Language	PPVT-III PLS-4 (Auditory Comprehension; Expressive Communication) CELF-4 (Concepts and following directions; Formulated sentences; Word structure; Word classes receptive; Word classes-expressive; Recalling sentences; Sentence structure; Expressive vocabulary) EOWPVT GFTA-2 Spontaneous language sample (choice according to child's age and time available for assessment)	YES Inter-rater agreement: 98.39–100%, except for GFTA-2 (66.23–94.81%)
2	Jessiman (2002)	mild receptive and expressive language delay (2)	6.2 (5.4–7) yrs	100%	CTA vs IPA (within subjects)	1. Speech 2. Language	CTA and IPA 1. SPAT-D IPA 2. – TOLD-P II - CELF- Preschool	YES Equivalence of scores
3	Manning et al. (2020)	typical development (62)	CTA: 24.87 (4.24) mo IPA: 24 (4.02) mo	nr	CTA vs IPA (between and within subjects)	Language	Language sampling (Mean length of utterance in morphemes; Number of different words used; Number of different words/total words; Number of errors; Number of omissions; Utterances audio signal intelligible; Utterances child speech intelligible)	YES No differences ($p > .05$)
4	Raman et al. (2019)	suspected of hearing, or speech or language or academic problems (15) typical development (17)	6.3 (.58) yrs	nr	CTA vs IPA (within subjects)	Language	ALD	YES No differences (receptive language: $z = 1.31$, $p > .05$; expressive language: $z = -1.09$, $p > .05$)

(*Continued*)

Table 1. (Continued).

	Article	Sample (n)	Mean age (sd/ age range)	% male	Experimental conditions	Cognitive function	Remote administered outcome measure	Agreement CTA and in person assessment/scoring
5	Sutherland et al. (2017)	history of reading difficulties and known/ suspected language impairment (23)	9.11 (8–12) yrs	78.26%	CTA vs in person scoring	1. Language 2. Behavior	1. CELF-4 (core subtests: Concepts and Following Directions, Word Structure/ Classes, Recalling Sentences, Formulated Sentences; two additional only in-person subtests: Understanding Paragraphs; Word Definitions) 2. Behavior rating scale (child's behavior during assessment: activity, interaction levels, attention to task, distractibility, type of distractors and anxiety)	YES 1. No trend in differences (95% limits of agreement) Inter rater reliability: r = .96–1.00, p < .01 Agreement in severity level: 96% 2. No differences (p > .05)
6	Sutherland et al. (2019)	ASD (13)	11.12 (9.5 – 12.3) yrs	76.92%	CTA vs in person scoring	1. Language 2. Behavior	1. CELF-4 (core subtests: Concepts and Following Directions, Word Classes, Recalling Sentences, Formulated Sentences; two additional only in-person subtests: Understanding Paragraphs; Word Definitions) 2. Behavior rating scale (child's behavior during assessment: activity, interaction levels, attention to task, distractibility, type of distractors and anxiety)	YES 1. No trend in differences (95% limits of agreement) Inter rater reliability: r = .94 – .99, p < .01 Agreement in severity level: 76.92% 2. No differences: p > .05
7	Waite et al. (2006)	speech disorder (6)	5.3 (4.3–6.8) yrs	nr	CTA vs in person scoring	Speech	SWAT Intelligibility in conversation (Likert scale 7 points) Oro motor structure and function (Dichotomous rating)	YES Level of agreement: 91–100%, except for double oral movements (63%) Inter and intra-rater agreement for online ratings: >70%
8	Waite et al. (2010a)	diagnosis of language impairment (12) difficulties in language (13)	7.2 (1.3) yrs;	68%	CTA vs in person scoring	Language	CELF-4 (Concepts and Following Directions, Word Structure, Recalling Sentences, Formulated Sentences)	YES No differences: p > .006 Level of agreement: k = .97-.99 Inter- and intra- rater reliability: ICC = .84–1.00

(*Continued*)

Table 1. (Continued).

	Article	Sample (n)	Mean age (sd/ age range)	% male	Experimental conditions	Cognitive function	Remote administered outcome measure	Agreement CTA and in person assessment/scoring
9	Waite et al. (2012)	speech disorders (20)	5 (4–9) yrs	65%	CTA vs in person scoring	Speech	1. Oro-motor structure and function (Likert scale; face at rest, four single oral movements, two double oral movements and three diadochokinetic tasks 2. Intelligibility in conversation (Likert scale 5 points)	YES 1. Level of agreement: >70%, K = .12-.74 Inter- and intra- rater reliability: ICC = .19–1.00 2. Level of agreement: >70% Inter- and intra- rater reliability: ICC = .64-.86
10	Ciccia et al. (2011)	typical development (411)	nr (up to 6) yrs	nr	1. CTA vs IPA (between subjects) 2. CTA vs IPA (within subjects)	1. Speech and Language 2. Hearing	1. – REEL-3 - SKOLD or PLS-4 abbreviated version - PLS-4 (articulation screener) 2. – Tympanometry (screening of middle ear health) - DPOAE (screening of hearing acuity)	YES 1. Reliability: 100% 2. Reliability: 84%
11	Werfel et al. (2021)	hearing loss (15)	nr	nr	CTA vs IPA (within subjects)	1.Speech and Language 2. Phonological memory 3.Working memory	1. nr 2. nr 3. nr, digit span	YES Reliability: .72–96, except for phonological memory and recording (–.10-.79)
12	Guiberson et al. (2015)	language impairment (34)typical development (48)	53.65 (8.84) mo	46.34%	CTA vs IPA (within subjects)	Verbal short-term memory	CTA: NWR IPA: SPLS-4 Additional: SDLQ for parents	YES Inter-rater agreement for online ratings: 97% Correlation with language measure: r = .55, p < .01

(*Continued*)

Table 1. (Continued).

	Article	Sample (n)	Mean age (sd/ age range)	% male	Experimental conditions	Cognitive function	Remote administered outcome measure	Agreement CTA and in person assessment/scoring
13	Kronenberger et al. (2021)	deafness (28) typical development (36)	14.7 (9–22) yrs	50%	CTA vs IPA (within subjects)	1. Verbal short-term memory 2.Working memory	CTA: 1. WJ-IV (Nonword Repetition) - WRAML-2 (Story Memory, Sentence Memory) 2. WISC-V (Digit Span Forward, Digit Span Backward, Letter–Number Sequencing) IPA: 1. – Test of Nonword Repetition - HINT-C 2. – Computerized Visual Digit Span test (forward, backward) - WISC–V (Letter–Number Sequencing)	YES 1. Correlations: r > .36, p < .05, except for nonword repetition (r = .33, p < .1 deaf group) and digit span backward (r = .26, p > .05 deaf group, r = .27, p > .05 typical group) 2. No differences (p > .05)
14	Harder et al. (2020)	demyelinating disorders (58)	13.10 (3.56) yrs	38%	CTA vs IPA (within subjects)	1. Learning and memory 2. Verbal abilities and working memory 3. Motor-free processing speed 4. Visuo-motor integration 5. Executive functions 6. Academic learning	1. CVLT-C, CVLT-II 2.WISC-V (Vocabulary, Digit Span Forward and Backward) 3. SDMT (Oral version) 4. VMI-6 5. D-KEFS (Verbal Fluency (letter and category); Color Word Interference (color naming, word reading, inhibition)) 6. WJ-III (Letter-Word Identification, Reading Fluency, Calculation, Math Fluency, Word Attack)	YES No differences: p = .42-.99
15	Ragbeer et al. (2016)	juvenile Batten disease (3) typical development (1)	14.3 (2.8) yrs	100%	CTA vs in person scoring	1. Verbal intelligence 2. Memory and learning 3. Executive functions	1. WISC-IV (Similarities, Vocabulary, Information, Digit Span) 2. WRAML-2 (Story Memory) 3. Verbal Fluency test	YES Level of agreement: >85%, except for verbal fluency (60%)
16	Hodge et al. (2019a)	SLD (reading) (23) SLD (reading) and ADHD (10)	9.11 (8–12) yrs	60.60%	CTA vs in person scoring	Intelligence	WISC-V	YES Correlation: r > .98, p < .01 Differences close to zero

(*Continued*)

Table 1. (Continued).

	Article	Sample (n)	Mean age (sd/ age range)	% male	Experimental conditions	Cognitive function	Remote administered outcome measure	Agreement CTA and in person assessment/scoring
17	Petrill et al. (2002)	typical development (52)	7.81 (.87) yrs	51.92%	CTA vs IPA (within subjects)	Intelligence	CTA: – WISC III: Similarities subtest, Vocabulary subtests, Symbol Search subtest, Picture Completion subtests, – adaptation of McCarthy Scales of Children's Abilities: Conceptual Grouping subtest – Nonword Repetition test short form IPA: SB Intelligence Scale short form(Pattern Analysis, Bead Memory, Quantitative, and Verbal subtests)	YES Correlation: r = .31-.64, p < .05
18	Wright (2020)	typical development (256)	nr (6–16) yrs	49,22% IPA, 54,69% CTA	CTA vs IPA (between subjects)	Intelligence	– WISC-V – KBIT–2 (only in person)	YES Differences within the threshold for equivalence, except for Letter-number sequencing Correlations (indices and IQ): ICC = [.1,+1]
19	Hamner et al. (2021)	Billing diagnosis of 893 patients: ADHD (546) Psychiatric disorders (261) Medical conditions (142) Other (93) *possible comorbidity*	10.1 (2.9) yrs	61%	CTA vs IPA (between subjects)	1. Intelligence 2. Academic abilities	– WISC-V (Similarities, Matrix Reasoning, Digit Span, Vocabulary, Visual Puzzles) – KTEA-3 (Letters and Words, Math Concepts)	YES No differences (p > .05), except for Visual Puzzles (B = .96, p = .01, d = .33) and Math Concepts (B = 2.95, p = .03, d = .18)

(*Continued*)

Table 1. (Continued).

	Article	Sample (n)	Mean age (sd/ age range)	% male	Experimental conditions	Cognitive function	Remote administered outcome measure	Agreement CTA and in person assessment/scoring
20	Hodge et al. (*Hodge, et al., 2019b)	SLD (reading) (27) SLD (reading) and ADHD (10)	10.1 (8–11) yrs	54.05%	CTA vs in person scoring	Reading skills	WRMT-III (Word Identification, Word Attack, Passage Comprehension) TOWRE-2 (Sight Word Efficiency, Phonemic Decoding Efficiency) MultiLit (MultiLit Sight Word Test, MultiLit Word Attack Test) DST	YES Correlation: r > .86, p < .01, except for letter-sound knowledge and blending skills (r = .79, p > .05) Differences close to zero
21	Waite et al. (2010b)	delays or suspected delays in reading and/or spelling (20)	9.11 (1.7) yrs	75%	CTA vs in person scoring	1. Phonological awareness 2. Spelling 3. Reading	1.QUIL (nonword spelling, nonword reading, syllable identification, syllable segmentation, spoken rhyme recognition, visual rhyme recognition, phoneme detection, and phoneme segmentation) 2. SAST (Form 1) 3. Neale-3 (Form 1)	YES Level of agreement: k = .92–1.00, except for the nonword reading (65%) and the reading error classification (75%) Inter- and intra- rater reliability: ICC = .89–1.00
22	Stain et al. (2011)	schizophrenia (5) schizoaffective disorder (2) psychosis not otherwise specified (3) major depressive disorder with psychotic features (1)	20 (14–27) yrs	45.45%	CTA vs IPA (within subjects)	1. Reading skills 2. Neuropsychological functions 3. Psychiatric symptoms 4. Quality of life 5. Social and Occupational Functioning	1. WTAR 2. – WMS-R (Logical Memory Subtest) - WAIS-III (Digit Span Subtest) - COWAT 3. BPRS 4. AQoL 5. SOFAS	YES Correlation: r = .81-.96, p < .01, except for digit span (r = .59, p > .05) Differences close to zero, except for WTAR
23	Salinas et al. (2020)	ADHD, ASD, language disorders, concussion, psychiatric diagnoses (67)	10.2 (2–18) yrs	55.22%	CTA vs IPA (within subjects)	1. – Communication and social interaction skills – Restricted and repetitive patterns of behavior, interests and/or activities 2. Executive functions 3. Adaptive behavior	1.- CARS-2 – ADOS-2 2. – NEPSY-II word generation subtest – D-KEFS (verbal fluency) 3. – Vineland-3 (comprehensive interview form) Firstly: child's interview, child's behavioral observation, clinical interview with parents	YES Concordance: 100%

Legend: **ADHD** (attention deficit hyperactivity disorder); **ADOS-2** (Autism Diagnostic Observation Schedule – Second Edition); **ALD** (Assessment of Language Development); **AQoL** (Assessment of Quality of Life); **ASD** (autism spectrum disorder); **VMI-6** (Beery-Buktenica Developmental Test of Visual Motor Integration, Sixth Edition); **BPRS** (Brief Psychiatric Rating Scale); **CARS-2** (Childhood autism rating scale 2nd edition); **CELF-Preschool** (Clinical Evaluation of Language Fundamentals Preschool); **CELF-4** (Clinical Evaluation of Language Fundamentals-4th edition); **COWAT** (Controlled Oral Word Association Test); **CTA** (Cognitive Tele-Assessment); **CVLT-C** (California Verbal Learning Test-Children's Version); **CVLT-II** (California Verbal Learning Test-Second Edition); **Dd** (Developmental delay); **D-KEFS** (Delis-Kaplan Executive Function System); **DPOAE** (Distortion Product Otoacoustic Emissions); **DSM 5** (Diagnostic and Statistical Manual of Mental Disorders fifth edition); **DST** (Dalwood Spelling Test); **EOWPVT** (Expressive One Word Picture Vocabulary Test); **GFTA-2** (Goldman Fristoe Test of Articulation-2nd edition); **HINT-C** (Hearing in Noise Test for Children);**KBIT-2** (Kaufman Brief Intelligence Test, Second Edition); **KTEA-3** (Kaufman Test of Educational Achievement, Third Edition); **IPA** (In person assessment); **MO** (months) **Neale-3** (Neale Analysis of Reading Ability, 3rd Edition); **NEPSY-II** (Developmental Neuropsychological Assessment-Second Edition); **nr** (not reported); **NWR** (Spanish nonword repetition); **PLS-4** (Preschool Language Scale-4th edition); **PPVT-III** (Peabody Picture Vocabulary Test-3rd edition); **QUIL** (Queensland University Inventory of Literacy); **REEL-3** (Receptive-Expressive Emergent Language Test – Third Edition); **SAST** (South Australian Spelling Test); **SB Intelligence scale** (Stanford-BinetIntelligence Scale); **SDLQ** (Spanish Preschool Language Scale); **SDMT** (Symbol Digit Modalities Test); **SKOLD** (Screening Kit of Language and Development); **SLD** (Specific Learning Disorders); **SOFAS** (Social and Occupational Functioning Assessment Scale); **SPAT-D** (Structured Photographic Articulation Test featuring – Dudsberry); **SPLS-4** (Spanish Preschool Language Scale, 4th edition); **STAT** (Screening Tool for Autism in Toddlers & Young Children); **SWAT** (Single Word Articulation Test); **TOLD-P III** (Test of Language Development Primary Third Edition); **TOWRE-2** (Test of Word Reading Efficiency-Second Edition); **Vineland-3** (Vineland Adaptive Behavior Scale, 3rd edition); **VMI-6** (Visual Motor Integration – 6 edition); **WAIS-III** (Wechsler Adult Intelligence Scale, Third Edition); **WISC III** (Wechsler Intelligence Scale for Children); **WISC-IV** (Wechsler Intelligence Scale for Children–Fourth Edition); **WISC-V** (Wechsler Intelligence Scale for Children, Fifth Edition); **WMS-R** (Wechsler Memory Scale-Revised); **WJ-III** (Woodcock-Johnson, Third Edition); **WJ-IV** (Woodcock-Johnson Tests of Cognitive Ability);**WRAML-2** (Wide Range Assessment of Memory and Learning–SecondEdition); **WRMT-III** (Woodcock Reading Mastery Tests-Third Edition); **WTAR** (Wechsler Test of Adult Reading); **YRS** (years).

Methodological and procedural characteristics

Across 23 studies, 2193 children were assessed with CTA on different cognitive domains: language, communication and speech, learning and memory, intelligence, visuo-motor integration, executive functions and academic abilities.

The population involved across and within the selected studies is very heterogeneous as children belonged to the following groups: typical development (n = 8; Ciccia et al., 2011; Guiberson et al., 2015; Kronenberger et al., 2021; Manning et al., 2020; Petrill et al., 2002; Ragbeer et al., 2016; Raman et al., 2019; Wright, 2020), speech or language disorders or difficulties (n = 9; Eriks-Brophy et al., 2008; Guiberson et al., 2015; Jessiman, 2002; Raman et al., 2019; Salinas et al., 2020; Sutherland et al., 2017; Waite et al., 2006, 2010a, 2012), autism spectrum disorder (n = 2; Salinas et al., 2020; Sutherland et al., 2019), attention and hyperactivity disorders (n = 4; Hamner et al., 2021; Hodge et al., 2019a, *Hodge, et al., 2019b; Salinas et al., 2020), learning disorders or difficulties (n = 5; Hodge et al., 2019a, *Hodge, et al., 2019b; Raman et al., 2019; Sutherland et al., 2017; Waite et al., 2010b), psychiatric disorders (n = 3; Hamner et al., 2021; Salinas et al., 2020; Stain et al., 2011), and medical conditions (n = 6; Hamner et al., 2021; Harder et al., 2020; Kronenberger et al., 2021; Ragbeer et al., 2016; Salinas et al., 2020; Werfel et al., 2021).

The sample age of the children who participated in the studies can be split into three ranges: preschoolers (0–6 years) (n = 9; Ciccia et al., 2011; Eriks-Brophy et al., 2008; Guiberson et al., 2015; Hamner et al., 2021; Manning et al., 2020; Salinas et al., 2020; Waite et al., 2006, 2012; Werfel et al., 2021), schoolers (6–12 years) (n = 18; Eriks-Brophy et al., 2008; Hamner et al., 2021; Harder et al., 2020; Hodge et al., 2019a, *Hodge, et al., 2019b; Jessiman, 2002; Kronenberger et al., 2021; Petrill et al., 2002; Ragbeer et al., 2016; Raman et al., 2019; Salinas et al., 2020; Sutherland et al., 2017, 2019; Waite et al., 2006, 2010a, 2010b, 2012; Wright, 2020); adolescents (12–18 years) (n = 7; Hamner et al., 2021; Harder et al., 2020; Kronenberger et al., 2021; Ragbeer et al., 2016; Salinas et al., 2020; Stain et al., 2011; Wright, 2020).

The geographical origin of the populations is not equally distributed throughout the world: thirteen studies were conducted in America (Ciccia et al., 2011; Eriks-Brophy et al., 2008; Guiberson et al., 2015; Harder et al., 2020; Hamner et al., 2021; Jessiman, 2002; Kronenberger et al., 2021; Manning et al., 2020; Petrill et al., 2002; Ragbeer et al., 2016; Salinas et al., 2020; Wright, 2020; Werfel et al., 2021), nine studies in Australia (Hodge et al., 2019a, *Hodge, et al., 2019b; Stain et al., 2011; Sutherland et al., 2017, 2019; Waite et al., 2006, 2010a, 2010b, 2012) and one study in India (Raman et al., 2019). Six research groups (Ciccia et al., 2011; Eriks-Brophy et al., 2008; Guiberson et al., 2015; Jessiman, 2002; Stain et al., 2011; Sutherland et al., 2017) involved children from rural, unserved, and remote areas. Few studies have investigated the cultural and social background of the participants. Five studies involved children from different ethnic groups, such as Caucasian, Hispanic, America Indian, Asian, Black or African American (Guiberson et al., 2015; Hamner et al., 2021; Harder et al., 2020; Manning et al., 2020; Salinas et al., 2020). Only two studies quoted the cultural level of the children's families: in the study by Wright (2020) the parents had at least some college experiences while in the study by Guiberson et al. (2015) the average of parents' schooling was around 9.5 years. Family incomes differed across studies, as they included low-income families (Ciccia et al., 2011) and middle-income families (Kronenberger et al., 2021; Manning et al., 2020).

The cognitive functions measures were as follows: speech and language (n = 11; Ciccia et al., 2011; Eriks-Brophy et al., 2008; Jessiman, 2002; Manning et al., 2020; Raman et al., 2019; Sutherland et al., 2017, 2019; Waite et al., 2006, 2010a, 2012; Werfel et al., 2021), verbal short and long-term memory (n = 5; Guiberson et al., 2015; Harder et al., 2020; Kronenberger et al., 2021; Ragbeer et al., 2016; Werfel et al., 2021), intelligence (n = 5; Hamner et al., 2021; Hodge et al., 2019a; Petrill et al., 2002; Ragbeer et al., 2016; Wright, 2020), academic abilities (n = 4; Hamner et al., 2021; Harder et al., 2020; Hodge et al., *Hodge, et al., 2019b; Waite et al., 2010b), neuropsychological functions, such as memory, visuomotor integration and executive functions (n = 6; Harder et al., 2020; Kronenberger et al., 2021; Ragbeer et al., 2016; Salinas et al., 2020; Stain et al., 2011; Werfel et al., 2021), communication and social interaction skills, restricted and repetitive patterns of behaviors, interests and/or activities (n = 1; Salinas et al., 2020) and quality of life, psychiatric symptoms, social and occupational functioning (n = 1; Stain et al., 2011). One study conducted the entire diagnostic process at distance, including not only the administration of a battery of tests, but also parents' interview and questionnaires, child's behavior observation and final assessment feedback with parents (n = 1; Salinas et al., 2020).

Not all studies directly compared CTA with IPA as some of them used CTA procedures only and then compared remote vs in presence scoring. The former refers to a scoring procedure in which the operator remotely records the child's responses through a technological device and he/she is not in the same room with the child. Conversely, in the "inpresence scoring" the operator records and scores the responses while sitting next to the child, in the same room. Thirteen studies compared CTA with IPA by within-subjects (n = 9; Guiberson et al., 2015; Harder et al., 2020; Jessiman, 2002; Kronenberger et al., 2021; Petrill et al., 2002; Raman et al., 2019; Salinas et al., 2020; Stain et al., 2011; Werfel et al., 2021), between-subjects (n = 2; Hamner et al., 2021; Wright, 2020) or mixed subjects designs (n = 2; Ciccia et al., 2011; Manning et al., 2020). Ten studies conducted a single assessment that was simultaneously scored in presence and in remote (n = 10; Eriks-Brophy et al., 2008; Hodge et al., 2019a, *Hodge, et al., 2019b; Ragbeer et al., 2016; Sutherland et al., 2017, 2019; Waite et al., 2006, 2010a, 2010b, 2012). In the study, by Salinas and colleagues (Salinas et al., 2020), authors compared for some children the diagnosis elaborated from results of the tests carried out remotely with the diagnosis based on the administration of the same tests in presence.

Assessment settings varied across studies: children's home (n = 7; Hamner et al., 2021; Harder et al., 2020; Kronenberger et al., 2021; Petrill et al., 2002; Salinas et al., 2020; Werfel et al., 2021; Wright, 2020), special needs service (n = 13; Ciccia et al., 2011; Eriks-Brophy et al., 2008; Hodge et al., 2019a; Jessiman, 2002; Manning et al., 2020; Stain et al., 2011; Sutherland et al., 2017, 2019; Waite et al., 2006, 2010a, 2010b, 2012; Wright, 2020), hospital (n = 1; Hodge et al., *Hodge, et al., 2019b), school (n = 4; Guiberson et al., 2015; Hodge et al., *Hodge, et al., 2019b; Raman et al., 2019; Wright, 2020) or a hotel (n = 1; Ragbeer et al., 2016). Usually, before starting the assessment, the operators ensured that child would be in a quiet room, with minimal physical distractors or noise and that no extra-assessment objects would be present on the child's working table (Harder et al., 2020; Petrill et al., 2002; Salinas et al., 2020). In most of the studies, the child was not alone, as there was an operator who provided technical assistance with the computer technology, test equipment and computer positioning as well as behavioral support.

Conversely, in order to maintain identical procedures between CTA and IPA, in Stain et al.'s (2011) and Harder et al.'s (2020) studies the child was alone, although in the latter case parents were in the house.

The assessment was usually conducted by expert clinicians or experimenters but in seven of the selected studies also parents were involved in the assessment procedures by suggesting the measures most fitting with the child's functioning (Ciccia et al., 2011), preparing the setting (Manning et al., 2020; Petrill et al., 2002; Salinas et al., 2020), participating in the diagnostic interview (Salinas et al., 2020), playing with the child (Manning et al., 2020) and filling out questionnaires (Guiberson et al., 2015; Hamner et al., 2021; Kronenberger et al., 2021).

Not all the articles reported the average time required for administering the tests in remote, but those who recorded this data (n = 12; Ciccia et al., 2011; Harder et al., 2020; Hodge et al., 2019a, *Hodge, et al., 2019b; Kronenberger et al., 2021; Manning et al., 2020; Petrill et al., 2002; Raman et al., 2019; Salinas et al., 2020; Stain et al., 2011; Sutherland et al., 2017, 2019), reported that CTA requires an average of 60 minutes (the range was between 30 and 90 minutes) to administer one test, factoring time to familiarize children with the set up, brief support persons, and occasional delays caused by connectivity issues. Authors noted that time correlated with the number of tests used and that it was usually higher for CTA than for IPA.

In all the studies, the equipment for CTA involved the use of common devices and technologies, such as a laptop, one or two video cameras, microphones and presentation of stimuli digitally or in paper form, with verbal explanation by the remote operator. The exceptions were the studies by Petrill et al. (2002) who used the telephone, Guiberson et al. (2015) who used a tablet both for the test presentation and videoconference, Harder et al. (2020), Kronenberger et al. (2021), and Manning et al. (2020) who made participants choose their preferred device (computer or tablet). Because of the child's age or mobility difficulties or researchers' choice, sometimes the child could use a sheet of paper to write down his/her answers and then return it to the researcher (Harder et al., 2020). Not all the studies had a digital system that showed stimuli or recorded answers, so the remote operator had to show stimuli through his/her video camera, or a specific training was necessary to teach the in-person operator how take note of the child's responses; however, the answers that were missed by the online operator could not be retrieved (Petrill et al., 2002).

Further details of the CTA-IPA procedures used in the selected studies is reported in Appendix C.

CTA feasibility

In order to measure satisfaction and collect comments, in seven studies the participants and the caregivers were interviewed or asked to anonymously respond to some questions at the end of the assessment. Results from the reports showed that participants and caregivers often exhibited curiosity about the CTA system and evaluated it as comfortable, easily accessible, useful and time cost saving; many families reported high level of satisfaction and recommended a videoconference assessment for anyone who would need it (Ciccia et al., 2011; Harder et al., 2020; Hodge et al., 2019a, *Hodge, et al., 2019b; Kronenberger et al., 2021; Stain et al., 2011; Sutherland et al., 2017, 2019). The examiners confirmed positive engagement and attitude by children during CTA (Eriks-Brophy et al., 2008; Hodge et al., 2019a, *Hodge, et al., 2019b; Ragbeer et al., 2016; Sutherland et al., 2017, 2019; Waite et al., 2010a) as if the digital administration of tests could have promoted attention and motivation (Hodge et al., 2019a, *Hodge, et al., 2019b; Waite et al., 2010a). For instance, in Sutherland and colleagues' study (2019), authors rarely reported behavioral or attentional problems, hyperactivity, anxiety or need to stop the evaluation during CTA, and only a few children behaved better in the face-to-face condition. High levels in children's motivation were reported also when CTA was conducted by telephone; anyway, during this assessment mode it was difficult to control for interfering variables, such as possible help from people in the surrounding environment, and it was hard to determine how much the child was paying attention to the task (Petrill et al., 2002). Small breaks were always planned and allowed, as it happens in IPA. In Eriks-Brophy and colleagues' study (2008), younger children appeared shyer than older ones, but the authors believed that technology was not the main reason for this tendency, as they were shy also during the IPA condition. Moreover, in Kronenberger and colleagues' study (2021), children with cochlear implants reported lower satisfaction than typical children. No studies investigated the relationship between children's behavioral difficulties and the presence of technological problems during the assessment.

Psychologists and experts rated CTA positively as well; they were satisfied with the remote screening technology, and they had high levels of confidence in the ratings (Hodge et al., 2019a, *Hodge, et al., 2019b; Sutherland et al., 2017). In Hodge and colleagues' study (2019a), psychologists and children enjoyed CTA, except in two cases in which the assessment was interrupted by bandwidth problems. According to the experts, no differences in children's behaviors were observed between CTA and IPA.

CTA reliability

CTA reliability was studied through different statistical techniques, varying across the selected studies; most of the studies used the degree of agreement between examiners and the statistical correlations between CTA and IPA.

The description of the characteristics of the studies and the reliability results found in each of the selected studies are summarized in Table 1 and reported below.

Almost all studies confirmed a good agreement between the CTA and IPA conditions, and between the diagnosis conducted remotely and in presence, except for a few cases. Of notice, studies varied for several methodological factors. A synthesis of the results is reported below according to the function investigated.

Speech and language

On speech and language tasks, a low agreement was mainly found in the case of articulation tests (Eriks-Brophy et al., 2008; Jessiman, 2002; Waite et al., 2006, 2012), in measures requiring the clinician to judge the correctness of the sounds and words pronounced by the child (Waite et al., 2006, 2010a), in the cases of unfamiliar nonwords and when the judgment was related to minimal speech pairs (e.g., stop consonant; Jessiman, 2002). In particular, Eriks-Brophy et al. (2008) found that the inter-rater agreement between CTA and in person scoring ranged between 98.39% and 100% for all measures of language except for the test of speech articulation (GFTA-2),which agreement was on average 80.26% (ranging between66.23and94.81%). In Jessiman's study (2002) differences between CTA and IPA were found in measures of perception of specific sounds, but only when standard rather than lapel microphones were used. Manning et al. (2020) documented a high transcript reliability and an absence of statistical differences between in person and remote transcriptions for all language measures ($p > .05$). In Raman and colleagues' study (2019), no significant differences between IPA and CTA were found in both receptive ($z = 1.31$, $p = .19$) or expressive language measures ($z = -1.09$, $p = .28$); moreover, language scores were concordantly classified. Results from Sutherland and colleagues' studies (2017, 2019) showed no differences between in-person scoring and CTA scoring (95% limits of agreement for each comparison), high values of inter-rater reliability on all measures (r = .94; −1.00, $p < .01$), 96% (study 1) and 76.92% (study 2) of agreement on the severity level of language disorders. Waite et al. (2006) documented high levels of inter- and intra-rater agreement (>70%) for the online scoring of most of the measures and a range from 91 to 100% of agreement between the in person and remote scores across the speech measures, except for double oral movements (63%). In a subsequent study, Waite et al. (2010a) confirmed high values of agreement (k = .97; .99), high values of intra- and inter-rater reliability (ICC = .84; 1.00) and no significant differences between CTA and in-person scoring in all on-line language measures ($p > .006$). In Waite and colleagues' study (2012) higher variability was documented for the oro-motor skills, characterized by moderate to good levels of agreement (>70%) between the online and in-person raters (k = .12; .74) and variable scores of intra- and inter-rater reliability (ICC = .19; 1.00). Also for intelligibility, a moderate degree of agreement was confirmed (70%) and intra- and inter-rater reliability was moderate to high (ICC: .64; .86). Ciccia et al. (2011) found 100% of reliability between tele- and in-person assessment for the measures of articulation and receptive-expressive language, pure tone hearing and distortion product otoacoustic emissions, and 84% of reliability for tympanometric measures. High reliability (.72; .96) on language measures was also confirmed by Werfel and colleagues' study (2021).

Verbal short and long-term memory

Some studies measured verbal short- and long-term memory. Low reliability was found on phonological memory in Werfel and colleagues' study (2021), especially on timed measures (−.10 to .79). Guiberson et al. (2015) found 97% inter-rater agreement on the remote measures of phonological working memory (NWR) that, in turn, significantly correlated with standardized language assessment scores (SPLS-4) ($r = .55$, $p < .01$). In Kronen berg and colleagues' study (2021) strong correlations were found between CTA and IPA across memory measures ($r > .50$, $p < .01$); correlations were modest for nonword repetition in children with cochlear implants ($r = .33$, $p < .1$). Harder et al. (2020) did not find significant differences between in-person and remote conditions for any measure of verbal learning and memory (.42 < adjusted p value < .99). Ragbeer et al. (2016) documented more than 93% of agreement between remote and in-person examiners on verbal long-term memory.

Intelligence

On intelligence measures, Ragbeer et al. (2016) documented more than 85% of agreement between remote and in-person examiners. Hodge et al. (2019a) documented that the in-person scoring and tele-assessments were significantly correlated (r > .98, p < .0 l) and scoring differences between in-person and tele-assessments were close to zero across all measures. Accordingly, in Petrill and colleagues' study (2002) a significant correlation between remote and in-person assessments was found for several intelligence subtests (r-scores ranging between = .31; .64, p < .05). In Wright and colleagues' study (2020) all confidence intervals and the difference scores between confidence intervals for all subtest scores fell within the threshold for remote and in-person equivalence except that for the letter-number sequencing subtest, where higher scores were obtained in the in-person condition. Correlations for the in-presence indices and IQ did not significantly differ between remote and in person groups (ICC[−1;+1]). No significant differences (p > .05) were found between IPA and CTA scores by Hamner et al. (2021) in any of the intelligence measures, except for the visual puzzle task, in which the CTA group obtained higher scores, even if with a small effect size (d = .33).

Academic abilities

Studies measuring academic learning suggested a good CTA-IPA agreement as well. As mentioned previously, Harder and colleagues' study (2020) also measured academic abilities and did not find significant differences between the in-person and remote conditions. Hamner et al. (2021) did not find any differences between IPA and CTA scores in the letters and words subtest (p > .05), while a significant difference was found for the math concepts subtest, in which the CTA group obtained higher scores, even if with a small effect size (d = .18). In the second study by Hodge et al. (*Hodge, et al., 2019b) it was confirmed that the scoring differences between in-person and tele-assessments were close to zero for all measures, and that there was a significant correlation between in-person and tele-assessments (r > .86, p < .0 l) with the exception for letter-sound knowledge and blending skills (r = .79, p > .05). Waite et al. (2010b) found an inter-rater agreement close to 1 (k = .92; 1.00) for all scores, except for the nonword reading raw score in theQUIL test (65% of agreement) and the reading error classification in the Neale-3 test (75.9% of agreement); high intra- and inter-rater reliability was documented for online parameters (ICC = .89; 1.00). Stain et al. (2011) found a significant correlation between the remote and in-person assessment of reading (r = .93, p < .01), even though higher reading skills ratings for the remote than in the in-presence assessment were documented.

Neuropsychological functions

Concerning the tele-assessment of other neuropsychological functions, Harder et al. (2020) did not find differences between CTA and IPA for executive functions, verbal abilities, working memory, motor-free processing speed, visuo-motor integration. In Stain and colleagues' study (2011) a significant correlation between the remote and in person assessment was found for the logical memory subtest (r = .81,p > .01) and the control oral word association test (r = .96, p < .001), but not for the digit span subtest (r = .59, p > .01) even if differences between IPA and CTA were close to zero in all measures. Ragbeer et al. (2016) found a modest level of agreement for verbal fluency

(60%); however the agreement increased to 80% when children with typical development were excluded from the analysis. On working memory measures, Kronenberger and colleagues' study (2021) found strong correlations ($r > .41$; $p < .05$), except for the digit span backward task, and no differences ($p > .05$) between CTA and IPA in children with cochlear implants. High reliability was found by Werfel et al. (2021) on verbal working memory.

Other functions

A few articles using CTA to study behavior documented high levels of agreement. In Sutherland's studies (2017, 2019), results showed that ranks of children's behaviors did not differ between CTA and IPA in neither of the studies ($p > .05$). Salinas et al. (2020) found a 100% of agreement between the remote and in person classification of communicative and social behavior. In Stain's study (2011), the level of agreement for social and occupational functioning was 100% and the mean differences between in person and remote assessment were close to zero.

Technical characteristics

In 14 studies (Hamner et al., 2021; Harder et al., 2020; Hodge et al., 2019a, *Hodge, et al., 2019b; Jessiman, 2002; Kronenberger et al., 2021; Manning et al., 2020; Raman et al., 2019; Sutherland et al., 2017, 2019; Waite et al., 2006, 2010a, 2010b, 2012) technology problems and difficulties were reported. They were related to the following aspects: room setup (especially within the child's house), connectivity (temporary screen freezing, bandwidth congestion, loss of connection and consequent need to refresh connection, slow connections that could cause distortion in the audio signal and/or audio breakup), audio and video low quality (i.e., audio or video latency, lag, delay, echo which prevents from giving sudden prompts), volume control, child playing with the microphone or microphone cord causing audio difficulties, delay in the transition of pictures, lags in responses,need to restart the browser; need to set up the camera; lighting (especially when a child gave his/her answer on a stimulus book, online clinician could have some problems to correctly see him/her),touch screen (a good level of thumb pressure was required to record child's answer),low voice and audio volume,children's movement and position in front of the web camera, environmental distractors (people entering in the room, high level of noise) and distractions from the technical equipment (for example, computer size, microphone worn).

Although technical problems appeared to invalidate CTA in very few cases (1.2% Manning et al., 2020; Raman et al., 2019; 4.3% in Sutherland et al., 2017), they impacted on testing time and child's cooperation (Raman et al., 2019).

CTA always required some technical preparation for both remote and, when included, in person examiners (Ciccia et al., 2011; Eriks-Brophy et al., 2008; Harder et al., 2020; Hodge et al., 2019a, *Hodge, et al., 2019b; Jessiman, 2002; Ragbeer et al., 2016; Raman et al., 2019; Sutherland et al., 2017; Waite et al., 2010a, 2010b, 2012). Remote examiners needed to be adequately prepared for distant administration, as they had to guide the room setting and to monitor the assessment procedure or any problems that parents, facilitator or child could have. At the same time, when included, the in-presence examiner or facilitator needed to receive a good preparation in order to adjust the webcam and the technological tools, control the volume and the position of the child in front of the computer, reproduce words/instructions that the remote examiner did not understand well, bring the child's attention back to the video, and make test materials, such as booklets and blocks, available. The in-presence role could be carried by an expert operator but also by a family member or a teacher and it was especially needed with younger children and preschoolers or older children with attention difficulties (Ciccia et al., 2011; Manning et al., 2020). When specific videoconference systems were used, in-presence assistants did not need to be actively involved in the test administration because stimuli were presented through computer (Guiberson et al., 2015; Hodge et al., 2019a, *Hodge, et al., 2019b; Jessiman, 2002; Raman et al., 2019; Sutherland et al., 2017, 2019; Waite et al., 2006, 2010a, 2010b, 2012; Wright, 2020).

In the case of diagnostic procedures entirely conducted remotely, authors (Salinas et al., 2020) suggested the need to include at least three phases: first, collecting information from parents on the child's functioning and characteristics; second, observing, and evaluating the child through standardized tests; and finally, providing feedback to parents as it happens in IPA.

Both for remote assessment and remote diagnosis, besides technological problems, the most influencing limit of CTA that was reported across the studies was the lack of physical presence, which may have reduced the understanding of non-verbal and implicit communication. These characteristics decrease, in turn, the operator–child relationship and the use of reengaging strategies (e.g., mimic encouragement) often required in the assessment of children, especially in early ages (Raman et al., 2019). For example, in CTA the examiner and the child rarely looked directly into each other's eyes and the loss of such a communication could have potentially increased the child's difficulties in emotional, motivational and behavioral regulation.

A detailed description of good practices suggested by the selected studies to increase the validity of CTA is reported in Appendix D.

Discussion

Nowadays, Cognitive Tele-Assessment (CTA) is an essential tool for evaluating cognitive functioning both in adults and children. CTA requires the adaptation of the settings, materials and methods that are usually used in the In-Person Assessment (IPA) to the remote and technological context. While IPA uses mainly paper and pencil stimuli, CTA is often based on images shown on computer screens that can differ in terms of size, resolution and other audio-visual parameters that may affect the subjects' performances. Most importantly, in CTA face-to-face interactions are replaced by technology-mediated interactions, with possible consequences on the child's behavior and compliance during testing. Given these and other differences between CTA and traditional IPA, it is of paramount importance to identify the most appropriate methodologies and procedures and be aware of the reliability of CTA in comparison to IPA as they vary across tests and participant characteristics. While nowadays CTA procedures in adults benefit from long-lasting research, the studies using CTA in children are more recent, thus their methodological characteristics have not been systematically described, nor future directions for developmental researchers and clinicians have been highlighted. The present review aimed at synthesizing the main methodological and procedural characteristics of the studies using CTA in children and to compare CTA and IPA in terms of feasibility and reliability.

The main findings across the 23 reviewed studies are very encouraging as they all suggest good levels of participants' satisfaction in using CTA and a high agreement between CTA and IPA, thus supporting the feasibility and validity of CTA in children. Moreover, one study suggests that the entire diagnostic process can be conducted totally remotely, a procedure that could represent an important option in the panorama of the CTA with children.

Nevertheless, given the small number of the available studies and the heterogeneity in the methods used, the present review may help in highlighting strengths and weaknesses of the previous literature and provide new directions for future use of CTA, rather than drawing general conclusions.

In several studies CTA and IPA evaluations were carried out on the same subjects and research contexts and, sometimes, without the use of parallel versions of the same tests, or without scheduling long intervals between CTA and IPA sessions (Ciccia et al., 2011; Harder et al., 2020; Jessiman, 2002; Raman et al., 2019; Salinas et al., 2020; Stain et al., 2011). These characteristics may favor learning effects that, in turn, could increase the correlation and agreement values between CTA and IPA. Future studies comparing CTA and IPA need to use between-subject designs or parallel tests and longer inter-sessions intervals for within subjects designs.

Most of the studies concerned language skills (Ciccia et al., 2011; Eriks-Brophy et al., 2008; Jessiman, 2002; Manning et al., 2020; Raman et al., 2019; Sutherland et al., 2017, 2019; Waite et al., 2006, 2010a, 2012) while other skills, equally important for child development, such as reading, writing, math, attention, memory and executive functions, were much less investigated. This tendency may be in part ascribed to the fact that the tests of language are easier to transfer from the in person to the remote setting because they are mainly based on verbal interactions and visual materials displayed on a computer screen rather than on concrete objects or complex paradigms, such as those required by visuo-spatial or problem-solving tests. Batteries that rely on paper and pencil or physical tools require advanced preparation, as they may need to be sent to the children and their support persons ahead of the appointment time. The prevalence of studies using CTA for language rather than for other non-verbal cognitive functions could have clouded the results of the present review, for example, by skewing the CTA-IPA correlation toward either high or low values; nevertheless, the absence of CTA-IPA comparisons among different functions prevents any definitive conclusion.

The characteristics of the samples used are quite variable across the reviewed studies: some articles report results on a very small sample (e.g., Jessiman, 2002; Ragbeer et al., 2016; Waite et al., 2006), age ranges vary from early childhood to late adolescence, and populations may include children with atypical development as well as children with neuro developmental or physical disorders (Guiberson et al., 2015; Ragbeer et al., 2016; Raman et al., 2019). The age of the sample differed across studies also as a function of the cognitive function examined: language was mainly assessed in preschool children while neuropsychological functions, memory and intelligence in adolescents. Although no study investigated the effect of age on CTA reliability, it represents a factor to consider when choosing CTA. Furthermore, the material characteristics of the tests could interact with the population assessed: for example, younger children or children with physical limitations may find it difficult to participate in CTA, because some types of technologies require fine sensory and motor skills (e.g., use of the keyboard). On the one hand, such a high variability in the characteristics of the populations assessed suggests the usefulness of CTA in a wide range of children, but, on the other hand, it makes it difficult to generalize the reported findings to the whole developmental population or to provide insights on the different use of CTA across different ages and developmental conditions, also becasue of the low number of available studies.

All studies supported the conclusion that, to be effective, CTA must take into account several issues regarding the technological and examiners preparation. Precautions are especially needed when the function or the behavior investigated is device-dependent, as it is the case of language that may be disrupted by slow web connection or low-quality device (Eriks-Brophy et al., 2008; Waite et al., 2010a, 2010b). As per IPA, combining objective test data with questionnaires, interviews and other collateral information collected in person or by CTA represents a good clinical practice, especially for the remote assessment of language (Guiberson et al., 2015). The use of CTA to conduct an entire diagnostic process seems still premature; whenever the remote diagnostic path is not considered exhaustive according to clinicians or parents, the in person evaluation must always be included (Salinas et al., 2020). For this reason, it would be important to consider CTA as a useful adjunct to IPA and not a substitute for it. However, being able to conduct even just half of the assessment remotely would be cost- and time-saving, and it would allow to collect data from the child's environment (Salinas et al., 2020). Furthermore, CTA could be considered a valid tool for early screening of children's functioning before the symptoms reach high levels of dysfunctionality (Juárez et al., 2018).

An absolute strength of CTA, acknowledged across the studies, is the possibility of accessing geographic areas distant from universities, clinical centers and educational institutes, saving time, money and efforts for traveling (Ragbeer et al., 2016). Nevertheless, it must be considered that this advantage could be limited by the digital divide phenomena, that is the difference in accessibility to web and technological tools, that may still be present in several countries (European Parliament, 2015).

Moreover, CTA could support longitudinal research with samples representative of different geographical places by reaching also those children who have moved away from the research or clinical institute (Hodge et al., 2019a). Furthermore, experimenters and clinicians can reduce, by CTA, the time needed for the assessment path (take reservation, going to the clinical center, test administration, test scoring, feedback to parents), so that in some cases time results to be halved in comparison to standard assessment (Salinas et al., 2020). Notwithstanding the several advantages, it must be noted that CTA may require more time than IPA in the preparation of the logistic and technical parameters before starting assessment as it represents a delicate and complex phase. The degree of preparation could vary according to cognitive function assessed, the test used and the child's age. For example, a deeper preparation of the examiner could be needed to assess non-verbal cognitive functions requiring complex visuo-motor stimuli, and young children. Further, there can be barriers to the use of CTA: resistance by operators (Dunkley et al., 2010), belief that this type of assessment can only be conducted at the university and the potential challenges of establishing relationships with the child (Jessiman, 2002; May & Erikson, 2014).

Furthermore, for what concerns the limitations of CTA, assuming that all the mentioned proxy precautions are taken and that standardized data for remote assessment are available (Kline, 2015), CTA may still lack of several features, such as a genuine relationship between the examiner and the child (Raman et al., 2019), flexible procedures for administration and direct observation.The assessment procedure may be better controlled via the in person operator as a proxy. Although these aspects have not been systematically investigated, they should be taken into account. Furthermore, whenever CTA becomes highly integrated with IPA, it must be acknowledged that none of them can be the substitute of the other, as the two working contexts will remain different.

Limitations of the selected studies

Given that CTA in children has captured attention only in recent years, studies are low in number and highly variable in methods and objectives. Not all studies were aimed to compare CTA and IPA, thus providing data roughly answering to our research questions. Sometimes the samples included a very small number of children, mixed in terms of type of development (typical vs atypical), and could not guarantee an adequate statistical power. There may be a bias toward verbal functions, so that procedure and results for the tele-assessment of visuo-spatial cognitive functions are neglected. Furtherly, no studies could guarantee the representativeness of the population, and possible digital divide phenomena cannot be excluded. These and other characteristics represent a limit of the present review and prevents us from drawing general conclusions on CTA in children.

Further directions

Further studies answering to the limitations listed above may provide important data on the feasibility and reliability of CTA in children. An interesting perspective for future CTA studies is to increase the ICT implementation in CTA, for example, by adapting existing software (i.e., CANTAB and AWMA) for cognitive assessment in children or by converting paper-pencil tasks into digital forms. To do this, a teleconferencing platform would need to be embedded in the assessment software. On the other hand, future research should explore which tests cannot be converted into digital format and for which paper-and-pencil stimuli would have to be sent to the child and support person in advance. Moreover, more studies comparing different age ranges may contribute to increase our knowledge on the CTA reliability across developmental periods. Another important aspect to investigate in future studies is the public and clinician/operator perception of CTA: by understanding people's attitudes and opinions on CTA it will be possible to modify, adjust and support the tele-evaluation methods. In the studies cited in this review, it is possible that parents who gave their consent to CTA already had positive attitudes toward this assessment methodology, associated with the child's interest in computers and/or technologies. This last variable should be taken into consideration by future research too.

Finally, because of the higher probability of technical problems in CTA than in IPA, future studies must plan video recordings and systematic inter-rater scoring (Waite et al., 2006). How much technological issues in CTA, such as problems in internet connection and low video and audio quality, can negatively impact on the performance of children with behavioral difficulties must yet be investigated (Raman et al., 2019; Sutherland et al., 2017).

Conclusion

The reviewed studies suggest that the potential of CTA in children is promising. Results support the existence of a high correlation and agreement between CTA and IPA and suggest several good practices to be taken in order to conduct CTA properly. Nevertheless, remote assessment of cognitive child development cannot completely replace the traditional in person screening, assessment and treatment practices. It would be advisable to choose a methodology that is conceived and designed ad hoc for a specific user. Several features, such as child's age, level of attentiveness, engagement, compliance, physical or sensory or cognitive limitations, personal history, environment, culture, ethnic group, and geographical area could guide the examiner toward the choice of CTA tools. In fact, on the one hand, CTA enables greater reach of families who live geographically far from research and clinical centers, but on the other hand it is important, before offering a CTA, to investigate the cultural and ethnic background of the family, the preference to the use of technological tools and the actual presence and accessibility of these. This could be done through a short phone call. Moreover, some practical perspectives have to be considered before choosing CTA with children: the possibility for the child to have stable internet connection, access to a support person for the child, an appropriate setting in which the child should be assessed (for example, home, school, clinic,), whether or not a test has been previously used in the CTA, whether the copyrighted test has been digitalized for use in CTA, the reliability of the test administered via CTA, the need to send materials, back-up plans when CTA has to be abandoned and choice of video conference platform. Thus, the use of CTA should not be understood as a substitute for the face-to-face assessment but as a valid supplementary tool to in person assessment that may give a larger number of children access to educational and health services in a timely and economic way.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

ORCID

Costanza Ruffini http://orcid.org/0000-0001-8180-8965
Christian Tarchi http://orcid.org/0000-0003-4013-4794
Chiara Pecini http://orcid.org/0000-0001-6554-8366

References

*articles included in the systematic review

*Ciccia, A. H., Whitford, B., Krumm, M., & McNeal, K. (2011). Improving the access of young urban children to speech, language and hearing screening via telehealth. *Journal of Telemedicine and Telecare*, 17(5), 240–244. https://doi.org/10.1258/jtt.2011.100810

*Guiberson, M., Rodríguez, B. L., & Zajacova, A. (2015). Accuracy of telehealth-administered measures to screen language in Spanish-speaking preschoolers. *Telemedicine and e-Health*, 21 (9), 714–720. https://doi.org/10.1089/tmj.2014.0190

*Hamner, T., Salorio, C. F., Kalb, L., & Jacobson, L. A. (2021). Equivalency of in-person versus remote assessment: WISC-V and KTEA-3 performance in clinically referred children and adolescents. *Journal of the International Neuropsychological Society: JINS*, 1–10. Advance online publication. https://doi.org/10.1017/S1355617721001053

*Harder, L., Hernandez, A., Hague, C., Neumann, J., McCreary, M., Cullum, C. M., & Greenberg, B. (2020). Home-based pediatric teleneuropsychology: A validation study. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 35(8), 1266–1275. https://doi.org/10.1093/arclin/acaa070

*Hodge, M. A., Sutherland, R., Jeng, K., Bale, G., Batta, P., Cambridge, A., Detheridge, J., Drevensek, S., Edwards, L., Everett, M., Ganesalingam, C., Geier, P., Kass, C., Mathieson, S., McCabe, M., Micallef, K., Molomby, K., Pfeiffer, S., Pope, S., Tait, F., & Silove, N. (2019b). Literacy assessment via telepractice is comparable to face-to-face assessment in children with reading difficulties living in rural australia. *Telemedicine Journal and E-health: The Official Journal of the American Telemedicine Association*, 25(4), 279–287. https://doi.org/10.1089/tmj.2018.0049

*Hodge, M. A., Sutherland, R., Jeng, K., Bale, G., Batta, P., Cambridge, A., Detheridge, J., Drevensek, S., Edwards, L., Everett, M., Ganesalingam, K., Geier, P., Kass, C., Mathieson, S., McCabe, M., Micallef, K., Molomby, K., Ong, N., Pfeiffer, S., Pope, S., & Silove, N. (2019a). 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. https://doi.org/10.1177/1357633X18776095

*Jessiman, S. M. (2002). Speech and language services using telehealth technology in remote and underserviced areas. *Journal Of Speech Language Pathology And Audiology*, 27(1), 45–51.

*Kronenberger, W. G., Montgomery, C. J., Henning, S. C., Ditmars, A., Johnson, C. A., Herbert, C. J., & Pisoni, D. B. (2021). Remote assessment of verbal memory in youth with cochlear implants during the COVID-19 pandemic. *American Journal of Speech-Language Pathology*, 30(2), 740–747. https://doi.org/10.1044/2021_AJSLP-20-00276

*Manning, B. L., Harpole, A., Harriott, E. M., Postolowicz, K., & Norton, E. S. (2020). Taking language samples home: Feasibility, reliability, and validity of child language samples conducted remotely with video chat versus in-person. *Journal of Speech, Language, and Hearing Research: JSLHR*, 63(12), 3982–3990. https://doi.org/10.1044/2020_JSLHR-20-00202

*Petrill, S. A., Rempell, J., Oliver, B., & Plomin, R. (2002). Testing cognitive abilities by telephone in a sample of 6-to 8-year-olds. *Intelligence*, 30(4), 353–360. https://doi.org/10.1016/S0160-2896(02)00087-9

*Ragbeer, S. N., Augustine, E. F., Mink, J. W., Thatcher, A. R., Vierhile, A. E., & Adams, H. R. (2016). Remote assessment of cognitive function in juvenile neuronal ceroid lipofuscinosis (Batten disease): A pilot study of feasibility and reliability. *Journal of Child Neurology*, 31(4), 481–487. https://doi.org/10.1177/0883073815600863

*Raman, N., Nagarajan, R., Venkatesh, L., Monica, D. S., Ramkumar, V., & Krumm, M. (2019). School-based language screening among primary school children using telepractice: A feasibility study from India. *International Journal of Speech-language Pathology*, 21(4), 425–434. https://doi.org/10.1080/17549507.2018.1493142

*Salinas, C. M., Bordes Edgar, V., Berrios Siervo, G., & Bender, H. A. (2020). Transforming pediatric neuropsychology through video-based teleneuropsychology: An innovative private practice model pre-COVID-19. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 35(8), 1189–1195. https://doi.org/10.1093/arclin/acaa101

*Stain, H. J., Payne, K., Thienel, R., Michie, P., Carr, V., & Kelly, B. (2011). The feasibility of videoconferencing for neuropsychological assessments of rural youth experiencing early psychosis. *Journal of Telemedicine and Telecare*, 17(6), 328–331. https://doi.org/10.1258/jtt.2011.101015

*Sutherland, R., Trembath, D., Hodge, A., Drevensek, S., Lee, S., Silove, N., & Roberts, J. (2017). Telehealth language assessments using consumer grade equipment in rural and urban settings: Feasible, reliable and well tolerated. *Journal of Telemedicine and Telecare*, 23(1), 106–115. https://doi.org/10.1177/1357633X15623921

*Sutherland, R., Trembath, D., Hodge, M. A., Rose, V., & Roberts, J. (2019). Telehealth and autism: Are telehealth language assessments reliable and feasible for children with autism? *International Journal of Language & Communication Disorders*, 54(2), 281–291. https://doi.org/10.1111/1460-6984.12440

*Waite, M. C., Cahill, L. M., Theodoras, D. G., Busuttin, S., & Russell, T. G. (2006). A pilot study of online assessment of childhood speech disorders. *Journal of Telemedicine and Telecare*, 12 (3_suppl), 92–94. https://doi.org/10.1258/135763306779380048

*Waite, M. C., Theodoros, D. G., Russell, T. G., & Cahill, L. M. (2010a). Internet-based telehealth assessment of language using the CELF-4. *Language, Speech, and Hearing Services in Schools*, 41 (4), 445–458. https://doi.org/10.1044/0161-1461(2009/08-0131)

*Waite, M. C., Theodoros, D. G., Russell, T. G., & Cahill, L. M. (2010b). Assessment of children's literacy via an Internet-based telehealth system. *Telemedicine Journal and E-health: The Official Journal of the American Telemedicine Association*, 16(5), 564–575. https://doi.org/10.1089/tmj.2009.0161

*Waite, M. C., Theodoros, D. G., Russell, T. G., & Cahill, L. M. (2012). Assessing children's speech intelligibility and oral structures, and functions via an Internet-based telehealth system. *Journal of Telemedicine and Telecare*, 18(4), 198–203. https://doi.org/10.1258/jtt.2012.111116

*Werfel, K. L., Grey, B., Johnson, M., Brooks, M., Cooper, E., Reynolds, G., Deutchki, E., Vachio, M., & Lund, E. A. (2021). Transitioning speech-language assessment to a virtual environment: Lessons learned from the ELLA study. *Language, Speech, and Hearing Services in Schools*, 52(3) , 769–775. https://doi.org/10.1044/2021_LSHSS-20-00149

*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*, 32(9), 809–817. https://doi.org/10.1037/pas0000939

Alloway, T. P. (2007). *Automated Working Memory Assessment (AWMA)*. Harcourt Assessment.

American Psychological Association Services (APA Services). (2020b). *Guidance on psychological tele-assessment during the COVID-19 crisis*. https://www.apaservices.org/practice/reimbursement/health-codes/testing/tele-assessment-covid-19?fbclid=IwAR1d_YNXYS2Yc5mdIz_ZIYSkrrJ_6A9BQeKuIHxEEjjRh1XDR6fOYncM3b4

American Psychological Association. (2017). *Ethical principles of psychologists and code of conduct*. https://www.apa.org/ethics/code/index.aspx

American Psychological Association. (2020a) *Informed consent checklist for telepsychological services*. https://www.apa.org/practice/programs/dmhi/research-information/informed-consent-checklist

Corona, L. L., Weitlauf, A. S., Hine, J., Berman, A., Miceli, A., Nicholson, A., Stone, C., Broderick, N., Francis, S., Juárez, A. P., Vehorn, A., Wagner, L., & Warren, Z. (2020). Parent perceptions of caregiver-mediated telemedicine tools for assessing autism risk in toddlers. *Journal of Autism and Developmental Disorders*, 51(2), 476–486. https://doi.org/10.1007/s10803-020-04554-9

Cortiella, C., & Horowitz, S. H. (2014). *The state of learning disabilities: Facts, trends and emerging issues*. National Center for Learning Disabilities. www.ncld.org/wp-content/uploads/2014/11/2014-State-of-LD.pdf

Dunkley, C., Pattie, L., Wilson, L., & McAllister, L. (2010). A comparison of rural speech-language pathologists‘ and residents‘ access to and attitudes towards the use of technology for speech-language pathology service delivery. *International Journal of Speech-language Pathology*, 12(4), 333–343. https://doi.org/10.3109/17549500903456607

Edwards, M., Stredler-Brown, A., & Houston, K. T. (2012). Expanding use of telepractice in speech-language pathology and audiology. *Volta Review*, 112(3), 227–242. https://doi.org/10.17955/tvr.112.3.m.704

Eriks-Brophy, A., Quittenbaum, J., Anderson, D., & Nelson, T. (2008). Part of the problem or part of the solution? Communication assessments of Aboriginal children residing in remote communities using videoconferencing. *Clinical Linguistics & Phonetics*, 22(8), 589–609. https://doi.org/10.1080/02699200802221737

European Parliament. (2015, December). *Bridging the digital divide in the EU*. Briefing. https://www.europarl.europa.eu/RegData/etudes/BRIE/2015/573884/EPRS_BRI(2015)573884_EN.pdf

Farmer, R. L., McGill, R. J., Dombrowski, S. C., McClain, M. B., Harris, B., Lockwood, A. B., Powell, S. L., Pynn, C., Smith-Kellen, S., Loethen, E., Benson, N. F., & Stinnett, T. A. (2020). Teleassessment with children and adolescents during the coronavirus (COVID-19) pandemic and beyond: Practice and policy implications. *Professional Psychology, Research and Practice*, 51 (5), 477–487. https://doi.org/10.1037/pro0000349

Galusha-Glasscock, J. M., Horton, D. K., Weiner, M. F., & Cullum, C. M. (2016). Video teleconference administration of the repeatable battery for the assessment of neuropsychological status. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 31(1), 8–11. https://doi.org/10.1093/arclin/acv058

Green, R., Till, C., Al-Hakeem, H., Cribbie, R., Téllez-Rojo, M. M., Osorio, E., Hu, H., & Schnaas, L. (2019). Assessment of neuropsychological performance in Mexico City youth using the Cambridge Neuropsychological Test Automated Battery (CANTAB). *Journal of Clinical and Experimental Neuropsychology*, 41(3), 246–256. https://doi.org/10.1080/13803395.2018.1529229

Juárez, A. P., Weitlauf, A. S., Nicholson, A., Pasternak, A., Broderick, N., Hine, J., Stainbrook, J. A., & Warren, Z. (2018). Early identification of ASD through telemedicine: Potential value for underserved populations. *Journal of Autism and Developmental Disorders*, 48(8), 2601–2610. https://doi.org/10.1007/s10803-018-3524-y

Kline, P. (2015). *A handbook of test construction (psychology revivals): Introduction to psychometric design*. Routledge.

Kmet, L. M., Lee, R. C., & Cook, L. S. (2004). *Standard quality assessment criteria for evaluating primary research papers from a variety of fields*. Alberta Heritage Foundation For Medical Research.

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 Psychoeducational Assessment*, 38(8), 923–941. https://doi.org/10.1177/0734282920962839

Lezak, M. D. (1995). *Neuropsychological assessment: 3rd Edition*. Oxford Press.

Luciana, M., & Nelson, C. A. (2002). Assessment of neuropsychological function through use of the Cambridge Neuropsychological Testing Automated Battery: Performance in 4- to 12-year-old children. *Developmental Neuropsychology*, 22(3), 595–624. https://doi.org/10.1207/S15326942DN2203_3

May, J., & Erikson, S. (2014). Telehealth: Why not? Perspectives of speech pathologists not engaging in telehealth. *Journal of Clinical Practice in Speech-Language Pathology*, 16(3), 147–151.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Nagel, D. A., & Penner, J. L. (2016). Conceptualizing telehealth in nursing practice: Advancing a conceptual model to fill a virtual gap. *Journal of Holistic Nursing: Official Journal of the American Holistic Nurses' Association*, 34(1), 91–104. https://doi.org/10.1177/0898010115580236

O'Brian, S., Smith, K., & Onslow, M. (2014). Webcam delivery of the Lidcombe program for early stuttering: A phase I clinical trial. *Journal of Speech, Language, and Hearing Research: JSLHR*, 57 (3), 825–830. https://doi.org/10.1044/2014_JSLHR-S-13-0094

Pearl, P. L., Sable, C., Evans, S., Knight, J., Cunningham, P., Lotrecchiano, G. R., Gropman, A., Stuart, S., Glass, P., Conway, A., Ramadan, I., Paiva, T., Batshaw, M. L., & Packer, R. J. (2014). International telemedicine consultations for neurodevelopmental disabilities. *Telemedicine Journal and E-health: The Official Journal of the American Telemedicine Association*, 20(6), 559–562. https://doi.org/10.1089/tmj.2013.0275

Sherwood, A. R., & MacDonald, B. (2020). A teleneuropsychology consultation service model for children with neurodevelopmental and acquired disorders residing in rural state regions. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 35(8), 1196–1203. https://doi.org/10.1093/arclin/acaa099

Taylor, O. D., Armfield, N. R., Dodrill, P., & Smith, A. C. (2014). A review of the efficacy and effectiveness of using telehealth for paediatric speech and language assessment. *Journal of Telemedicine and Telecare*, 20(7), 405–412. https://doi.org/10.1177/1357633X14552388

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Guide for administering and scoring the fourth edition: Stanford–Binet intelligence scale*. Riverside.

Vallar, G., & Papagno, C. (Eds.). (2007). *Manuale di neuropsicologia: Clinica ed elementi di riabilitazione*. Il mulino.

Wadsworth, H. E., Dhima, K., Womack, K. B., Hart, J., Jr, Weiner, M. F., Hynan, L. S., & Cullum, C. M. (2018). Validity of teleneuropsychological assessment in older patients with cognitive disorders. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 33(8), 1040–1045. https://doi.org/10.1093/arclin/acx140

Appendix A. Keywords used in the three databases (Scopus, PsycInfo, PubMed)

SCOPUS

TITLE-ABS-KEY ((“pediatric*” OR “child*” OR “young children” OR “school-aged children” OR “youth”) AND (“tele-assessment” OR “telemedicine-based assessment” OR “teleneuropsychology” OR “videoconferencing” OR “in-person-based assessment” OR “remote assessment” OR “telehealth” OR “via telehealth” OR “telepractice” OR “telepsychology”) AND (“cognitive” OR “intelligence” OR “intellectual abilit*” OR “literacy” OR “math” OR “language” OR “speech” OR “memory” OR “learning” OR “attention” OR “perception” OR “visuo-spatial” OR “motor” OR “executive function*” OR “neuropsycholog*”) AND NOT (adult*)) AND NOT (animal*) AND (LIMIT-TO (LANGUAGE, “English”))

PSYCINFO

AB ((“pediatric*” or “child*” or “young children” or “school-aged children” or “youth”)) AND AB ((“tele-assessment” or “telemedicine-based assessment” or “teleneuropsychology” or “videoconferencing” or “in-person-based assessment” or “remote assessment” or “telehealth” or “via telehealth” or “telepractice” or “telepsychology”)) AND AB ((“cognitive” or “intelligence” or “intellectual abilit*” or “literacy” or “math” or “language” or “speech” or “memory” or “learning” or “attention” or “perception” or “visuo-spatial” or “motor” or “executive function*” or “neuropsycholog*”))

Filters: ENGLISH

PUBMED

(((“pediatric*”[Title/Abstract] OR “child*”[Title/Abstract] OR “young children”[Title/Abstract] OR “school-aged children”[Title/Abstract] OR “youth”)[Title/Abstract]) AND ((“tele-assessment”[Title/Abstract] OR “telemedicine-based assessment”[Title/Abstract] OR “teleneuropsychology”[Title/Abstract] OR “videoconferencing”[Title/Abstract] OR “in-person-based assessment”[Title/Abstract] OR “remote assessment”[Title/Abstract] OR “telehealth”[Title/Abstract] OR “via telehealth”[Title/Abstract] OR “telepractice”[Title/Abstract] OR “telepsychology”)[Title/Abstract])) AND ((“cognitive”[Title/Abstract] OR “intelligence”[Title/Abstract] OR “intellectual abilit*”[Title/Abstract] OR “literacy”[Title/Abstract] OR “math”[Title/Abstract] OR “language”[Title/Abstract] OR “speech”[Title/Abstract] OR “memory”[Title/Abstract] OR “learning”[Title/Abstract] OR “attention”[Title/Abstract] OR “perception”[Title/Abstract] OR “visuo-spatial”[Title/Abstract] OR “motor”[Title/Abstract] OR “executive function*”[Title/Abstract] OR “neuropsycholog*”)[Title/Abstract])

Filters: English, human, child: birth-18 years

Appendix B. Quality scores of the studies according to the checklist for quantitative studies developed by Kmet et al. (2004). Scores range from 0 to 1 (low: <.60; medium: from .61 to .80; high: from .81 to 1)

		1. Question/objective sufficiently described?	2. Study design evident and appropriate?	3. Method of subject/comparison group selection or source information/input variables described and appropriate?	4. Subject (and comparison group, if applicable) characteristics sufficiently described?	5. If interventional and random allocation was possible, was it described?	6. If intervention and blinding of investigators was possible, was it reported?	7. If interventional and blinding of subjects was possible, was it reported?	8. Outcome and (if applicable) exposure measure(s) well defined and robust to measurement/misclassification bias? Means of assessment reported?	9. Sample size appropriate?	10. Analytic methods described/justified and appropriate?	11. Some estimate of variance is reported for the main results?	12. Controlled for confounding?	13. Results reported in sufficient detail?	14. Conclusions supported by the results?	Score
1	Eriks-Brophy et al. (2008)	2	2	2	1	n/a	0	0	2	0	0	2	1	2	2	0.62
2	Jessiman (2002)	2	2	2	2	n/a	0	0	2	0	0	0	0	2	2	0.54
3	Manning et al. (2020)	2	2	2	2	n/a	0	0	2	2	2	2	1	2	2	0.81
4	Raman et al. (2019)	2	2	2	2	0	0	0	2	2	2	2	1	2	2	0.75
5	Sutherland et al. (2017)	2	2	2	2	n/a	0	0	2	2	2	2	1	2	2	0.81
6	Sutherland et al. (2019)	2	2	2	2	n/a	0	0	2	1	2	2	1	2	2	0.77
7	Waite et al. (2006)	2	2	2	2	n/a	0	0	2	0	0	2	1	2	2	0.65
8	Waite et al. (2010a)	2	2	2	2	2	0	0	2	2	2	2	2	2	2	0.86

(*Continued*)

Appendix B. (Continued).

		1. Question/objective sufficiently described?	2. Study design evident and appropriate?	3. Method of subject/comparison group selection or source information/input variables described and appropriate?	4. Subject (and comparison group, if applicable) characteristics sufficiently described?	5. If interventional and random allocation was possible, was it described?	6. If intervention and blinding of investigators was possible, was it reported?	7. If interventional and blinding of subjects was possible, was it reported?	8. Outcome and (if applicable) exposure measure(s) well defined and robust to measurement/misclassification bias? Means of assessment reported?	9. Sample size appropriate?	10. Analytic methods described/justified and appropriate?	11. Some estimate of variance is reported for the main results?	12. Controlled for confounding?	13. Results reported in sufficient detail?	14. Conclusions supported by the results?	Score
9	Waite et al. (2012)	2	2	2	2	2	0	0	2	2	2	2	2	2	2	0.86
10	Ciccia et al. (2011)	2	2	1	0	0	0	0	2	2	0	0	1	1	2	0.46
11	Werfel et al. (2021)	2	2	2	0	n/a	0	0	0	1	2	0	1	1	2	0.5
12	Guiberson et al. (2015)	2	2	2	2	n/a	0	0	2	2	2	2	1	2	2	0.81
13	Kronenberger et al. (2021)	2	2	2	2	n/a	0	0	2	2	2	2	1	2	2	0.81
14	Harder et al. (2020)	2	2	2	2	2	0	0	2	2	2	2	2	2	2	0.86
15	Ragbeer et al. (2016)	2	2	2	2	2	0	0	2	0	0	2	1	2	2	0.68
16	Hodge et al. (2019a)	2	2	2	2	n/a	0	0	2	2	2	2	1	2	2	0.81
17	Petrill et al. (2002)	2	2	2	2	n/a	0	0	2	2	2	2	1	2	2	0.81

(*Continued*)

Appendix B. (Continued).

		1. Question/objective sufficiently described?	2. Study design evident and appropriate?	3. Method of subject/comparison group selection or source information/input variables described and appropriate?	4. Subject (and comparison group, if applicable) characteristics sufficiently described?	5. If interventional and random allocation was possible, was it described?	6. If intervention and blinding of investigators was possible, was it reported?	7. If interventional and blinding of subjects was possible, was it reported?	8. Outcome and (if applicable) exposure measure(s) well defined and robust to measurement/misclassification bias? Means of assessment reported?	9. Sample size appropriate?	10. Analytic methods described/justified and appropriate?	11. Some estimate of variance is reported for the main results?	12. Controlled for confounding?	13. Results reported in sufficient detail?	14. Conclusions supported by the results?	Score
18	Wright (2020)	2	2	2	2	2	0	0	2	2	2	2	2	2	2	0.86
19	Hamner et al. (2021)	2	2	2	2	n/a	0	0	2	2	2	2	1	2	2	0.81
20	Hodge et al. (*Hodge, et al., 2019b)	2	2	2	2	n/a	0	0	2	2	2	2	1	2	2	0.81
21	Waite et al. (2010b)	2	2	2	2	2	0	0	2	2	2	2	2	2	2	0.86
22	Stain et al. (2011)	2	2	2	2	0	0	0	2	0	2	2	1	2	2	0.68
23	Salinas et al., 2020	2	2	2	2	n/a	0	0	2	2	0	0	1	2	2	0.65

Appendix C. Further details of the CTA-IPA procedures used by the selected studies

A brief description of the CTA-IPA procedures used across the studies is reported below.

In Eriks-Brophy et al. (2008) a speech-language pathologist remotely administered and scored the tests while another speech-language pathologist was in the room with the child and simultaneously scored the tests as soon as they were administered. A third researcher compared the two scores.

In Jessiman (2002) study a speech-language pathologist remotely administered one test (SPAT-D), through the Regional Satellite Based Telehealth System, while the child, his/her parents and a facilitator were in a remote site. After 3 days from the CTA, a speech language pathologist conducted the in person assessment of the articulation (SPAT-D) and phonological skills (TOLD-P II; CELF- Preschool) within the clinic setting. The language articulation re-assessment (SPAT-D) was conducted 3 months later by a remote speech-language therapist.

In Manning et al. (2020) CTA was conducted via video-chat at home and recorded by remote operators; all video-recordings were transcribed by undergraduate research assistants using the Systematic Analysis of Language Transcripts software (SALT). 38 video-recordings were transcribed twice by the same research assistance to assess coding reliability. In CTA conditions parents used home web-software available on their own devices. To evaluate the relation between CTA and IPA, authors compared video-recordings scoring between the two conditions.

In Raman and colleagues' (2019) study assessments were carried out in two conditions: in the IPA condition a speech language pathologist administered language screening tools at school, while in the CTA condition a remote speech language pathologist, connected from the hospital, assessed the child who was with a teaching assistant at school. Digitized stimuli were presented on a laptop using PowerPoint for both IPA and CTA; in this latter case an external webcam and stereo headsets were linked to the remote station, by a mobile hotspot/plugin dongle. Independent assessments were conducted during the same day: half of the sample was firstly assessed in person, whereas the other half was firstly assessed remotely.

In Sutherland and colleagues' (2017; 2019) studies a remote speech-language pathologist was connected by a metropolitan telehealth site and administered four language tests to a child located in another building together with an in person speech pathologist. In study 2 a parent was present during the assessment. Remote and in-presence clinicians scored performances simultaneously. Moreover, the in person speech pathologist administered two other tests before or after telehealth sessions.

In Waite et al. (2006)'s an in person speech pathologist and an online speech pathologist connected through eREHAB (Internet-based tele-health system), evaluated child performances. The remote clinician presented stimuli while the in person one rated the performances. After the evaluation session, the remote clinician completed ratings from the videorecordings. Recordings were scored again four weeks after (intra-rater agreement) and were also scored by an independent rater (inter-raters agreement).

In subsequent studies (Waite et al., 2010a, 2010b, 2012), Waite and colleagues used a methodology similar to Waite et al., 2015, in which remote and in person speech-language pathologists simultaneously scored different child language and speech skills. Test administration was conducted in two randomized conditions: remote-led condition (i.e., the remote clinician guided assessment, while the in person colleague merely observed) and in person-led condition (i.e., assessment directed by the in person clinician, while the other colleague merely observed remotely). To measure intra-rater reliability, clinicians rated recordings twice, after 4 weeks, and to study inter-rater reliability, an independent rater scored the recordings.

Ciccia et al. (2011) used a facilitator, an in person student in speech-language pathology or audiology, and a remote clinician from the clinical institute. The facilitator, who was supervised via videoconferencing on Skype, administered the hearing measures and sent results to the remote clinician for scoring. No further details about the procedure used for the remote assessing of language were provided. After about one year, the facilitator re-screened in presence some children for hearing and evaluated language skills of some children who did not previously receive remote assessment.

Werfel et al. (2021) conducted a CTA on 15 out of 23 children who received an in-person assessment the month before COVID-19. Authors used for CTA the same materials as IPA after adapting them for digital administration; they selected Zoom as a video conferencing platform and they used an app that measures sound level.

Guiberson et al. (2015) used a hybrid tele-health approach that implemented synchronous videoconferencing, videocasting, and traditional pen and paper tests and questionnaires. Children were with their parents in a room, while a bilingual English-Spanish researcher was connected with the child with a teleconferencing application on the iPad in a separate room. The remote researcher conducted the phonological working memory test by videoconference; in a subsequent session, children were remotely recorded during the narrative task and parents completed the questionnaires. Finally, other researchers administered the standardized language tests (SPLS-4) in a traditional face-to-face setting. 10% of the online assessments were video-recorded in order to establish inter-rater reliability of online test scoring.

In Kronenberg and colleagues' study (2021) a CTA on verbal short-term memory and working memory was conducted 1.6 years after the IPA. During CTA, a speech-language pathologist connected with the children over the Zoom Health platform. The examiner administered tests in standardized spoken format with his/her face in full view on children's computers.

Harder et al. (2020) conducted a test-retest study on both in person and tele-assessment conditions. Time between test and retest ranged between 1 to 50 days. The rater could be a primary trained examiner, a neuropsychologist or a psychometrist. For remote assessment children joined the remote rater through VSeeTM, a system for video conference that can be used on computer or tablet devices.

In Ragbeer et al. (2016) two randomized assessment's conditions were used: remote- led condition and in person-led condition. Remote rater used a platform for HD video conferencing (Vidyo) and a secure web-based audio-video link to join the child.

Hodge et al. (2019a, *Hodge, et al., 2019b) used a remote psychologist who interacted and administered the tests through Coviu, a web-based platform, while an in person acting as facilitator (a psychologist or a teacher) was next to the child in the community hub, in the school or in the hospital context. In person and remote operators scored performances simultaneously.

In Petrill and colleagues (2002) study, a research assistance reached the child at home to administer an intelligence scale (SB short form, Thorndike et al., 1986), subsequently the family received by mail the material for the administration the other intelligence scale (WISC III) that was subsequently conducted remotely by telephone.

In Wright (2020) study children were firstly tested inpresence with and intelligence scale (KBIT-2) and subsequently, they were tested according to their randomly assigned format, in remote or inpresence, with another intelligence scale (WISC V). The same examiner administered both tests. In the remote condition, the child and the psychologist were connected by a tele-health platform and a non-trained proctor was with the child. In order to determine that the administration of WISC V was equivalent across conditions (inpresence and remote), the Two-One-Sided-Tests was used (Lakens et al., 2017) according to which a confidence interval of 90% at $p < .05$ was established.

Hamner et al. (2021)studied intelligence and academic abilities comparing performances between a group of children and adolescents who received an in person assessment (608) in the period November 2019-March 2020 and a group who was involved in a remote assessment (285), using Q-Global platform, in the period April 2020-August 2020.

In Stain and colleagues' study (2011), in-presence and remote conditions were counterbalanced within the sample, with an inter-session interval up to 2 weeks; remote assessment was conducted by videoconferencing.

In Salinas et al. (2020) study a remote operator, located in the clinic, assessed the child who was at home with parents. Remote assessment was conducted by a virtual diagnostic tool. A follow-up was conducted for the in-person assessment on the 1.5% of the sample, constituted by children with diagnosis of autism spectrum disorder.

Appendix D. Good practices and guidelines for conducting CTA in children

In order to properly conduct CTA in children, several studies suggested the following good practices.

Room: the room must have good light and acoustics, being located within a silent area without environmental noise and with high signal quality. In the case of home-based assessment, it may be difficult for families to find a quiet space within the house; thus, examiners can remotely help parents to organize the room and to take precautions (Harder et al., 2020; Manning et al., 2020).

Equipment. Before making an appointment, especially when CTA is organized in the child's home, clinicians should ensure that adequate technological devices are available. In order to avoid that the use of tele-evaluations amplifies inequalities in society (e.g., in Harder et al., 2020 study almost a quarter of participants did not have access to necessary tools), it is necessary to ensure that all selected participants can satisfy the technological requirements in terms of devices, screen size, microphones, bandwidth.

Regardless of the location, it is suggested to use a handy and economical technology (for example, a portable speaker, a laptop instead of a monitor), that can be portable, adaptable to the child's need, and fit to different socio-economic conditions (Hodge et al., 2019a). A solution could be to equip the house or the school with the necessary tools, using technologies that are easily available in daily life, or to request the loan of material from nearby institutional centers. By using CTA in schools, it is possible to reach a considerable number of children at the same time and to guarantee everyone access to the research or screening process, in accordance with principles of social equalities.

Before starting the testing session, the remote assessor should investigate the child's behavior through a survey, preferentially with questions written on the computer screen rather than orally presented (Stain et al., 2011). Examiners need to monitor the child's behavior through a video camera to have eye contact with the child, in order to facilitate the relationship before starting the assessment, and to use a split screen technology to simultaneously view the stimuli and the child.

The measures of the headset microphone, table height and motor skills required by the device have to be adequate to the child's age and height (Waite et al., 2010a). The microphone should be easily worn and should be cordless to avoid tripping or being entangled (Jessiman, 2002).

If a mobile phone is used, although not recommended, it must be horizontally oriented, and the microphone must not be covered by other stuff.

The use of Wi-Fi or ethernet and the check of connectivity before each session is recommended (Manning et al., 2020); it is preferred to have a dedicated broadband cable connection, an increase of bandwidth for multiple users or to have several internet options available (e.g., mobile connectivity, LAN, Wi-Fi).

Child's familiarization with the examiner and the equipment. Children need to be prepared for distance assessment; it may be used a story (Sutherland et al., 2019), telling them what they will do, what tool (e.g., computer, microphone) they will use and where they will be during evaluation (e.g., laboratory, house's room). Furthermore, prior to evaluation, the child needs time to familiarize with the devices and the room (Eriks-Brophy et al., 2008). When the assessment is conducted in the house, the examiner can instruct parents to prepare the child to the CTA setting. Such a procedure will allow to avoid surprises or child's impulsive behaviors.

Moreover, the remote examiner needs to establish a relationship with the child prior to administering the tests; some information gathered on the child, a few simple questions to make the child feel at ease and listened to, allow for greater engagement on the part of the child.

When children show or have a history of attentional and motivational difficulties, the CTA procedures should be conducted with special precautions. The lack of physical presence prevents indeed the understanding of non-verbal and implicit communication or the use of reengaging strategies such as the eye contact between the examiner and the child (Raman et al., 2019). A way to overcome behavioral problems is to have an in-presence examiner who may control the setting and help the child and the parents on the need; such a role is particularly important when remote assessment is conducted with young children.

Family engagement. Especially with preschoolers, parents are actors in the CTA as they actively participate and involve the child in some tests. Having the possibility to remotely evaluate the child while interacting with the parents is undoubtedly a big advantage (Manning et al., 2020). Very often, in fact, in person assessments, especially at a behavioral level, underestimate the child's problematic behavioral manifestations, which is why it is always necessary to obtain information from the context of daily life (e.g., by parents or teachers). Furthermore, the presence of a familiar person, rather than a stranger, during the assessment allows the child to feel at ease and to show the best of her/his everyday character.

Examiner preparation. In-presence operators (i.e., research assistant, clinician, parents, teacher, caregiver) should participate in a short, but substantial session to receive a minimal training in order to know how handle the testing environment (e.g., positioning the child, monitoring environmental noise levels, placing the camera, adjusting lighting, setting up the room and general troubleshooting, etc.), technology (e.g., logging into application or videoconferencing or system, switching on laptop, connecting to internet, placing headphones, troubleshooting when loss in connectivity), child behavior (e.g., ensuring child's cooperation, boosting child's interest, etc.) and documenting the factors influencing the screening (Sutherland et al., 2017, p. 2019; Raman et al., 2019). When the in-presence facilitator is asked to operate within the child-remote operator dyad, it is necessary that he/she receives further preparation, to understand the activities and their implementation. This can be done through one or more preliminary meetings between remote examiner and facilitator but also, for example, through the dissemination of explanatory videos. Remote operator and in person facilitator should establish a relationship prior to the assessment, in order to coordinate and monitor any problem that can happen during the evaluation or to follow similar rules for scoring and to decide what in-presence operator can do or cannot do (Eriks-Brophy et al., 2008). The facilitator also plays the role of observing the child's behavior, movements and levels of interaction and eventually repeating the answers of the child to ensure the correct scoring by the remote assessor. If the facilitators are not expert, such as school assistants or parents, they should also receive training to adequately interact with the child according to his/her temperament and behavior (Raman et al., 2019).

Psychometric properties. In interpreting results and scoring performances, IPA standardization cannot be used for CTA. Conducting CTA requires psychologists to be aware of several recommendations, including those regarding the factors that can affect reliability and validity, but also psychometric data and their limitations, as well as cultural, ethical, and safety considerations (Eriks-Brophy, 2008).

Procedural Protocols. Especially for research studies, in order to guarantee objectivity and replicability of CTA, it would be useful to define a specific protocol which includes guidelines about digital platform requirements, examiner training and use of an informed in-presence operator (Wright, 2020).

Legal and ethical issues (American Psychological Association [APA], 2017; APA, 2020a; American Psychological Association Services (APA Services), 2020b). A specific informed consent including CTA procedure is required (Harder et al., 2020). When CTA uses tests that have not previously been validated, no conclusion can be drawn. In choosing the technological system, it is important to pay attention to the level of security guaranteed and the respect of sensitive data and privacy (e.g., using a low volume not audible by third parties, not disseminate screenshots or videos of the evaluation session etc., Hodge et al., *Hodge, et al., 2019b).