Comparing remote *versus* in-person assessment of learning skills in children with specific learning disabilities

Valentina Lampis[1,2], Chiara Dondena[1] , Chiara Mauri[1], Martina Villa[3,4,5], Antonio Salandi[1], Massimo Molteni[1], Chiara Cantiani[1] and Sara Mascheretti[1,2]

Abstract

Background: Interactive telemedicine applications have been progressively introduced in the assessment of cognitive and literacy skills. However, there is still a lack of research focusing on the validity of this methodology for the neuropsychological assessment of children with Specific Learning Disorder (SLD).

Methods: Seventy-nine children including 40 typically developing children (18 males, age 11.5 ± 1.06) and 39 children with SLD (24 males, age 12.3 ± 1.28) were recruited. Each participant underwent the same neuropsychological battery assessing reading accuracy, speed, and comprehension, writing, numerical processing, computation, and semantic numerical sense, twice (once during an in-person session (I) and once during a remote (R) home-based videoconference session). Four groups were subsequently defined based on the administration order. Repeated-measure-ANOVAs with assessment type (R vs. I testing) as within-subject factor and diagnosis (SLD vs. TR) and administration order (R-I vs. I-R) as between-subject factors, and between-group t-tests comparing the two assessment types within each time of administration, were run.

Results: No differences emerged between I and R assessments of reading accuracy and speed, numerical processing, and computation; on the contrary, potential biases against R assessment emerged when evaluating skills in writing, reading comprehension, and semantic numerical sense. However, regardless of the assessment type, the scores obtained with I and R assessments within the same administration time point overlapped.

Discussion: These results partially support the validity and reliability of the assessment of children's learning skills *via* a remote home-based videoconferencing system. Implementing telemedicine as an assessment tool may increase timely access to primary health care and to support research activity.

Keywords

Interactive telemedicine, remote home-based videoconference assessment, learning skills, specific learning disorders, validity and reliability

[1]Child Psychopathology Unit, Scientific Institute, IRCCS Eugenio Medea, Bosisio Parini, Italy
[2]Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy
[3]Department of Psychological Sciences, University of Connecticut, Storrs, CT, USA
[4]The Connecticut Institute for Brain and Cognitive Sciences, University of Connecticut, Storrs, CT, USA
[5]Yale Child Study Center Language Sciences Consortium, New Haven, CT, USA

Corresponding authors:
Chiara Cantiani, Child Psychopathology Unit, Scientific Institute IRCCS Eugenio Medea, Bosisio Parini, Lecco, Italy.
Email: chiara.cantiani@lanostrafamiglia.it

Sara Mascheretti, Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy; Child Psychopathology Unit, Scientific Institute IRCCS Eugenio Medea, Bosisio Parini, Lecco, Italy.
Email: sara.mascheretti@unipv.it

Introduction

Specific learning disorders (SLDs) are complex neurodevelopmental disorders characterized by difficulties in learning and using academic skills in spite of adequate neurological and sensorial functioning, educational opportunities, and average intelligence.[1] About 5–15% of school-age children across different languages and cultures are affected by these disorders which are often associated with negative secondary functional and psychosocial outcomes.[1] The clinical diagnosis of such disorders requires psychometric evidence from a battery of individually administered and culturally appropriate tests of academic skills that are norm-referenced or criterion-referenced.[1]

In the last few decades, healthcare services have been subjected to a gradual digitalization process,[2] and interactive telemedicine applications have been progressively introduced in clinical and educational practices.[3–7] The term "interactive telemedicine" (ITM hereafter) refers to a real-time interaction between patient and clinician conducted remotely via a videoconferencing system. The use of ITM tools has notably increased for both assessment of neuropsychological functioning[8–12] and remediation treatments.[13–16] Further interest in ITM has recently arisen due to the physical distancing imposed by the SARS-CoV-2 pandemic that forced, for a protracted period, the interruption or substitution of traditional face-to-face assessments in both research activity and clinical practice.[5,12,17–19] The advantages of ITM are multiple and well-documented. They include simpler and more rapid access to diagnostic services, especially for people who live in rural or remote areas and those with cultural, socio-economic, psychological, or physical disadvantages; increased rates of early diagnosis and, consequently, early interventions.[2,20–22] Yet, there is still much resistance to the use of neuropsychological assessment through ITM, especially in developmental age.[23]

Although the videoconference neuropsychological assessment in children is assumed to be reliable and valid for research purposes, its use in clinical settings is still to be ascertained. The settings and procedural rules for diagnostic screening through ITM are not well-defined, and there is no consensus among clinicians about how to use these tools in diagnostic protocols.[3] Moreover, it is still not clear if the lack of a physical relationship between the examiner could substantially affect the feasibility, acceptability, and validity of ITM tools by children.[24–27]

Previous studies reported good reliability for ITM assessment of cognitive and literacy skills.[6,7,28–29] Waite and colleagues[28] showed good validity and reliability of an internet-based videoconference system for the assessment of literacy skills in a sample of 20 children aged 8–13 years. By comparing the scoring assigned by the experimenter present at the administration with those of a remotely connected rater, Hodge and colleagues[6,7] found high correlations between teleassessment and face-to-face-rated scores obtained by children aged 8–12 years with reading difficulties. In a large cross-sectional study including a total of 893 clinically referred children and adolescents aged 4–17 years assessed via either teleassessment or traditional in-person testing, Hamner and colleagues[29] found equivalency across methods of examination delivery without clinically meaningful differences in scores of cognitive and academic tests. Although these studies provided evidence about the reliability of ITM assessment of cognitive and literacy tests, they did not lead to the validation of these remote videoconference neuropsychological assessments. Recently, Harder and colleagues[22] validated home-based teleneuropsychological assessment in a pediatric sample of 58 participants aged 6–20 years. Each participant was administered the same brief neuropsychological battery of common measures twice, once during an in-person session and once during a remote home-based videoconference session. Order of sessions was counterbalanced and time between assessments ranged from 1 to 50 days. Results showed no significant differences in scores obtained during the two different assessment types irrespective of the order of administration.

Technology is nowadays commonly used in everyday life for adults and children[30]; the use of ITM for neuropsychological assessment of cognitive and literacy skills in children with SLD can therefore represent a valid alternative in clinical and research practice. In the present study, we aimed to test the validity and reliability of a remote home-based videoconference assessment of learning-related skills (i.e., reading accuracy and speed, reading comprehension, writing, and mathematics) in both children with SLD and typically developing (TD) children attending primary and middle school by administering the same neuropsychological battery twice, once during an in-person session and once during a remote home-based videoconference session. The significance of this study is threefold: (a) the inclusion of both TD children and children with a clinical diagnosis of SLD allowed us to investigate whether the remote home-based videoconference assessment was reliable and valid regardless of the diagnosis; (b) testing each subject with the same neuropsychological battery twice (once during an in-person session and once during a remote home-based videoconference session) allowed us to evaluate and disentangle the effect of administration order, assessment type, and diagnosis; and, (c) since the same investigator administered and corrected both the remote home-based videoconference and in-person tests of each child, we can exclude possible biases deriving from interrater variability.

Material and methods

Sample

The study sample consists of 79 children including 40 TD children (18 boys, mean age 11.6 ± 1.1) and 39 children

with a clinical diagnosis of SLD (24 boys, mean age 12.25 ± 1.3). Subjects with a clinical diagnosis of SLD[1] were recruited at the Child Psychopathology Unit, Scientific Institute, IRCCS Eugenio Medea (Bosisio Parini, Lecco) or by recalling participants from previous research projects.[31] In particular, we included children with SLD who received a clinical diagnosis of developmental dyslexia and/or developmental dysorthography and/or developmental dyscalculia. TD were recruited among students attending primary and middle school. In order to be included in the present study, TD must not have any previous clinical diagnosis of neurodevelopmental disorders and any previous clinical-referred sensorial and/or neurological deficits. Both children with SLD and TD had average performance IQ (i.e., >85 as measured by either the Cattel test[32] or the matrix subtest of the Wechsler Intelligence Scale for Children-Fourth Edition[33]) were native Italian speakers and attended school from Grade 5 to Grade 8 (from 10 to 14 years old). Informed written consent was obtained by parents of all the participants included in the study.

Each participant underwent the same neuropsychological assessment twice, once during a traditional in-person (I) session and once during a remote (R) videoconference session in a quiet room in their own home (hereafter, assessment type). Time between assessments ranged from 2 to 10 days (M = 6.75, SD = 1.74 days); the same experimenter administered and corrected the tests of the same child in both sessions.

Regarding subjects with SLD, children entering the clinical evaluation at the Child Psychopathology Unit, Scientific Institute, IRCCS Eugenio Medea (Bosisio Parini, Lecco) were assigned to the I-R administration order (SLD_I-R; n = 20); children recalled from previous research projects investigating the genetic underpinnings of SLD[31] were assigned to the R-I administration order (SLD_R-I; n = 19). Regarding this latter group, children were recalled if they participated in previous studies at least 12 months before the data collection for the current project. No significant differences in diagnosis of SLD ($\chi^2_{(3)}$ = 5.292, p = 0.152) and in clinically assessed comorbidity with other neurodevelopmental disorders ($\chi^2_{(2)}$ = 4.256, p = 0.119) emerged between the SLD groups (SLD_R-I vs. SLD_I-R). TD were randomly assigned to R-I (n = 20) or I-R administration (n = 20) order. The four groups (SLD_I-R, SLD_R-I, TD_I-R, and TD_R-I) significantly differed for age ($F_{3,78}$ = 6.569, p < 0.001), but they did not significantly differ for sex ($\chi^2_{(3)}$ = 3.276, p = 0.351) and performance IQ ($F_{3,75}$ = 2.069, p = 0.112).

Neuropsychological assessment

For both assessment types, subjects underwent the following neuropsychological evaluation:

1. *Reading* as assessed by passage-of-text,[34] four lists of 24 single unrelated words and three lists of 16 single unrelated pseudowords.[35] Regarding passage-of-text reading, texts increase in complexity with grade level. For all tests, children were asked to read as fast and accurately as possible; accuracy (number of errors) and speed (time in seconds) were recorded, and z-scores were obtained on grounds of grade norms from the general population. Text reading tests were standardized on more than 12,000 Italian-speaking children attending from Grade1 to Grade8.[34] T-tests comparing children with developmental dyslexia and typical readers confirmed the discriminant validity of all texts in all grades (p < 0.001) in both speed and accuracy parameters; reliability was at best moderate in both accuracy (test–retest rs = 0.59–0.79) and speed (test–retest rs = 0.84–0.96).[34] Single unrelated words and pseudowords reading tests were standardized on 929 Italian-speaking children attending from Grade2 to Grade8; both tests showed good reliability for accuracy (test–retest r = 0.56) and speed (test–retest r = 0.77).[35] Since r correlations among the different tests within accuracy and speed were all substantial regardless of assessment type (I: mean rs = 0.64 and rs = 0.83 for accuracy and speed, respectively; R: mean rs = 0.64 and rs = 0.84 for accuracy and speed, respectively), two composite scores (i.e., Reading Accuracy and Reading Speed) were obtained by averaging the z-scores from the three scores within each parameter (i.e., accuracy and speed);

2. *Reading comprehension* as assessed by the reading comprehension test.[34] Children were asked to read the text (aloud or in their mind) and then answer multiple choice questions. Texts increase in complexity with grade level. Reading comprehension tests were standardized on more than 12,000 Italian-speaking children attending from grade 1 to grade 8.[34] T-tests comparing children with developmental dyslexia and typical readers confirmed the discriminant validity of all texts in almost all grades (Grade3: p = 0.073, Grade4: p < 0.150, Grade5: p < 0.001, Grade6: p < 0.001, Grade7: p = 0.073, and Grade8: p < 0.001)[34]; reliability was at best modest (test–retest rs = 0.29–0.79).[34] Accuracy (number of correct responses) was recorded, and z-scores were obtained on grounds of grade norms from the general population;

3. *Writing* as assessed by the untimed sentences writing under dictation subtest.[35] Children were asked to write down 12 sentences containing homophones. The experimenter dictated the sentences following the child's writing rhythm. Writing test was standardized on 929 Italian-speaking children attending from Grade2 to Grade8; reliability was at best modest (test–retest r = 0.37).[35] Number of errors was recorded, and z-scores were obtained on grounds of grade norms from the general population;

4. Mathematical abilities as assessed by the BDE 2—Batteria discalculia evolutiva, Test per la diagnosi dei

disturbi dell'elaborazione numerica e del calcolo in età evolutiva—8–13 anni.[36] The battery was standardized on more than 800 Italian-speaking students attending from Grade3 to Grade8. An exploratory factor analysis identified a 3-factor structure (numerical processing, computation, and semantic numerical skills), explaining 42.28% of the total variance; the internal consistency of the three factors, measured with Cronbach's alpha, was satisfactory (0.74 for numerical processing, 0.62 for computation, and 0.69 for semantic numerical skills). We therefore assessed the following mathematical skills:

(a) *Numerical processing* as assessed by three subtests: (1) counting: children were asked to count to 80–140 as fast and accurately as they could and the examiner recorded the time (in seconds). Subsequently, children were asked to count backward from 140 as much as they could within the time they took to count forward; (2) number reading: children were asked to read as fast and accurate as possible four lists of 48 numbers within 60 s; and, (3) number writing under dictation subtests: children were asked to repeat and write down 18 numbers;

(b) *Computation* as evaluated by two subtests: (1) multiplication: children were asked to solve as fast and accurately as possible 18 multiplications read by the experimenter within 3 s; and, (2) mental calculation: children were asked to solve as fast and accurately as possible 18 mental calculations (both additions and subtractions) read by the experimenter within 30 s;

(c) *Semantic numerical skills* as tested by two subtests: (1) number triplets: children were asked to identify the highest number within each triplet ($n = 18$) and to complete as many trials as they could in 2 min; and, (2) approximate calculations: children were given multiple choice (four possible answers are presented) and were asked to select the correct answer for up to 18 operations within 2 min.

(d) For all subtests, accuracy (number of correct items) was recorded, and z-scores were obtained on grounds of grade norms from the general population.

Online implementation of the neuropsychological assessment

PsychoPy builder version 2021.1.4[37] was used to convert the neuropsychological assessment into the remote home-based videoconference assessment type. Although the implementation of the testing battery from I to R assessment type was accomplished by keeping the stimuli and the answering procedure as similar as possible, some minor changes were necessary for some tests (see the "Changes in the tests' administration during the two assessment

types). PsychoPy is an open-source package for running experiments in Python. In addition, it is possible to run the same experiments online through an open-source repository, the Pavlovia platform,[38] by compiling a JavaScript version of the task. On the Pavlovia platform, each experiment can be accessed via the corresponding URL without being able to download any stimulus. Two URLs were created: one for reading (with passage-of-text reading and reading comprehension specific to the grade attended by the subject) and one for mathematics. Each subtest was preceded by written instructions that were also read aloud by the experimenter. In order to proceed during testing, participants were instructed to press the spacebar whenever they were ready. Answers provided by the participants through the mouse/touchpad or spacebar were stored on the Pavlovia platform and subsequently downloaded for scoring. Unlike other platforms for the administration of digital psychological testing, the Pavlovia platform does not require participants to download any software on their devices, is General data protection regulation-GDPR compliant, and has a relatively low monetary cost.[38] The R assessment type was performed during a Zoom videoconference during which the experimenter shared the URLs with participants and explained general instructions for running the experiment. Each Zoom videoconference was recorded. Regardless of the device used and its graphics card, stimuli's visualization on the screen was unchanged.

Changes in the tests' administration during the two assessment types

Reading accuracy and speed. For all the reading subtests, all the psychophysical features (i.e., font, font size, line spacing) of the print version were maintained in the R assessment.

In both assessment types, reading accuracy was evaluated by the experimenter who recorded errors while the child was reading and the unrelated single words and pseudowords reading lists were presented one at a time. Reading speed was tracked by the experimenter using a chronometer in the I assessment, whereas it was automatically recorded by the software (i.e., a chronometer started automatically when the stimulus appeared on screen and children were instructed to press the spacebar as soon as they finished reading to stop it) in the R assessment. However, in order to avoid a potential loss of information in case the child forgot to press the spacebar, the experimenter simultaneously used a chronometer to manually record speed.

Reading comprehension. In both assessment types, subjects were asked to read a passage of text. During the answering phase, in the I assessment, text and multiple-choice questions were presented on different papers and subjects could go back

to the text as many times as they wanted, follow their own order of response and correct the given answers at any time using a pencil. While in the R assessment, the text remained displayed on the left side of the screen and one question at a time was presented on the right side; subjects needed to press the spacebar to see next questions, had to pick among the multiple options using the mouse/touchpad and, due to technical restrictions, after providing the answer, they could not turn back to correct it.

Writing. The instructions did not change between the two assessment types except that, in the R assessment, subjects were asked to show the paper to the experimenter as soon as the administration ended and to send a copy of it *via* email at the end of the neuropsychological assessment.

Numerical processing. For all the subtests within this neuropsychological domain, the administration did not change between the two assessment types except for the number reading and number writing under dictation subtests. In particular, in the R assessment of the number reading subtest, the lists of numbers disappeared after 2 min. Regarding the number writing under dictation subtest, children were asked to show the paper to the examiner as soon as the administration ended and to send a copy of it *via* email at the end of the neuropsychological assessment.

Computation. In both assessment types, the experimenter recorded the answers provided by the children. In the R assessment, the stimuli were previously recorded, uploaded, and then played *via* the Pavlovia platform. Children were instructed to press the spacebar as soon as they were ready to move on to the subsequent item. In order to assess if the response provided by the child was given within the expected maximum time, a sound was automatically played by the software after 3 s (for multiplications) and 30 s (for mental calculations).

Semantic numerical skills. In the I assessment, all items and their multiple choices were visible on the paper at the same time so that subjects could follow their own order of response and correct the given answers at any time using a pencil. On the contrary, in the R assessment, one item at the time appeared on the screen. Children had to pick among the multiple options using the mouse/touchpad and, after providing the answer, subjects could not review or correct it.

Statistical analyses

A set of repeated-measure ANOVAs were run separately for each neuropsychological domain (Reading Accuracy, Reading Speed, Reading Comprehension, Writing, Numerical processing, Computation, Semantic Numerical skills) with assessment type (R vs. I testing) as within-subject factor, diagnosis (SLD vs. TD), and administration

order (R-I vs. I-R) as between-subject factors, and age as covariate. Then, in order to control for potential practice effects and following significant interactions emerging from the overall ANOVA, we conducted separate between-group t-tests comparing the two assessment types within each time of administration (comparing R-I vs. I-R groups in first evaluation and in second evaluation). All the statistical analyses were run in SPSS Version 28.0.[39]

Results

Table 1 showed the descriptive statistics for each neuropsychological domain (Reading Accuracy, Reading Speed, Reading Comprehension, Writing, Numerical processing, Computation, Semantic Numerical skills) in each assessment type (i.e., I vs. R) within each group (SLD_I-R, SLD_R-I, TD_I-R, TD_R-I).

Reading accuracy

A main effect of diagnosis was found ($F_{1,74} = 44.210$, $p < 0.001$, $\eta^2 = .374$) suggesting that children with SLD had overall lower scores than TD (SLD: M = −1.462, SE = .183; TD: M = .291, SE = .181). Moreover, the interaction between assessment type and administration order ($F_{1,74} = 17.529$, $p < 0.001$, $\eta^2 = .192$) was significant, and the interaction between assessment type, diagnosis, and administration order showed a trend toward significance ($F_{1,74} = 3.772$, $p = 0.056$, $\eta^2 = .049$). Within the two administration orders (I-R and R-I), the difference between the two assessment types was significant (I-R: $p < 0.001$, R-I: $p = 0.016$), suggesting the presence of a practice effect. Interestingly, this effect seemed to be modulated by the diagnosis (TD_I-R: $p = 0.084$, TD_R-I: $p = 0.786$, SLD_I-R: $p = 0.006$, SLD_R-I: $p = 0.002$). In other words, only children with SLD in both administration order groups had better scores in the second assessment than in the first assessment regardless of the assessment type. On the contrary, TD in both administration order groups showed similar performances in the first and the second assessments (see Figure 1).

In order to control for practice effect and the significant diagnosis effect, four separate between-group t-tests were used. Results showed that there were no significant differences between the two assessment types within each time of administration (children with SLD at time 1: $t_{(37)} = −.200$, $p = 0.843$; children with SLD at time 2: $t_{(37)} = −.306$, $p = 0.762$; TD at time 1: $t_{(38)} = −.389$, $p = 0.699$; TD at time 2: $t_{(38)} = .774$, $p = 0.444$). These findings suggest that, within each timepoint, the two different assessment types did not lead to any difference in the subjects' scores.

Reading speed

A main effect of diagnosis was found ($F_{1,74} = 65.620$, $p < 0.001$, $\eta^2 = .470$) suggesting that children with SLD had

Table 1. Descriptive statistics for each neuropsychological domain in each assessment type within each group (mean and standard deviation are reported).

	SLD_R-I	SLD_I-R	TD_R-I	TD_I-R
Reading Speed_I	−1.56 ± 0.98	−1.97 ± 1.65	0.88 ± 0.79	−0.00 ± 0.82
Reading Accuracy_I	−1.15 ± 1.23	−1.78 ± 1.78	0.31 ± 0.66	0.16 ± 0.93
Reading Comprehension_I	−0.45 ± 0.97	−0.28 ± 1.39	0.46 ± 0.94	0.06 ± 0.92
Writing_I	−1.18 ± 1.43	−1.45 ± 2.1	0.03 ± 0.62	−0.05 ± 0.72
Numerical Processing_I	−0.58 ± 0.58	−0.69 ± 1.13	0.70 ± 0.57	0.12 ± 0.54
Computation_I	−1.47 ± 1.00	−1.32 ± 1.12	0.32 ± 0.60	−0.12 ± 0.90
Semantic Numerical Sense_I	−0.58 ± 0.67	−1.05 ± 1.08	0.25 ± 0.48	−0.05 ± 0.74
Reading Speed_R	−2.03 ± 1.02	−1.57 ± 1.55	0.49 ± 0.78	0.28 ± 0.83
Reading Accuracy_R	−1.68 ± 1.31	−1.28 ± 1.54	0.27 ± 0.69	0.46 ± 0.60
Reading Comprehension_R	−0.88 ± 0.82	−0.13 ± 0.9	0.07 ± 0.91	0.19 ± 1.04
Writing_R	−1.74 ± 1.84	−1.47 ± 2.40	0.16 ± 0.63	−0.08 ± 0.86
Numerical Processing_R	−0.77 ± 0.73	−0.56 ± 0.84	0.36 ± 0.54	0.40 ± 0.65
Computation_R	−1.63 ± 1.12	−1.07 ± 1.17	0.39 ± 0.57	0.06 ± 0.78
Semantic Numerical Sense_R	−1.00 ± 0.84	−0.81 ± 0.92	−0.12 ± 0.75	0.05 ± 0.57

I: In-person assessment; R: Remote videoconference assessment.

overall lower scores than TD (SLD: M = −1.722, SE = .177; TD: M = .350, SE = .175). Moreover, the interaction between assessment type and administration order was significant ($F_{1,74} = 90.031$, $p < 0.001$, $\eta^2 = .549$). Within the two administration orders (I-R and R-I), the difference between the two assessment types was significant (R-I: $p < 0.001$ and I-R: $p < 0.001$), suggesting the presence of a practice effect (see Figure 2).

In order to control for practice effect, two separate between-group t-tests were used. Results showed that, regardless of the diagnosis, there were no significant differences between the two assessment types within each time of administration (assessment at time 1, $t_{(77)} = −.701$, $p = 0.486$ and assessment at time 2, $t_{(77)} = −.978$, $p = 0.331$). These findings suggested that, within each timepoint, the two different assessment types did not lead to any difference in the subjects' scores.

Reading comprehension

A main effect of diagnosis was found ($F_{1,73} = 8.180$, $p = 0.006$, $\eta^2 = .101$) suggesting that children with SLD had overall lower scores than TD (SLD: M = −.425, SE = .151; TD: M = .192, SE = .147). Moreover, the interaction

between assessment type and administration order was significant ($F_{1,73} = 6.934$, $p = 0.010$, $\eta^2 = .087$). Within the two administration orders (I-R and R-I), the difference between the two assessment types was significant only in the R-I order ($p = 0.007$) and not in the I-R order ($p = 0.345$). These results suggest that a practice effect was present only in the R-I order (see Figure 3). In order to control for practice effect, two separate between-group t-tests were used. Results showed that, regardless of the diagnosis, there were no significant differences between the two assessment types within each time of administration (assessment at time 1, $t_{(77)} = 1.156$, $p = 0.251$ and assessment at time 2, $t_{(76)} = −.007$, $p = 0.995$). These findings suggested that, within each timepoint, the two different assessment types did not lead to any difference in the subjects' scores.

Writing

A main effect of diagnosis was found ($F_{1,71} = 16.025$, $p < 0.001$, $\eta^2 = .184$) suggesting that children with SLD had overall lower scores than TD (SLD: M = −1.411, SE = .245; TD: M = −0.016, SE = .238). In addition, the
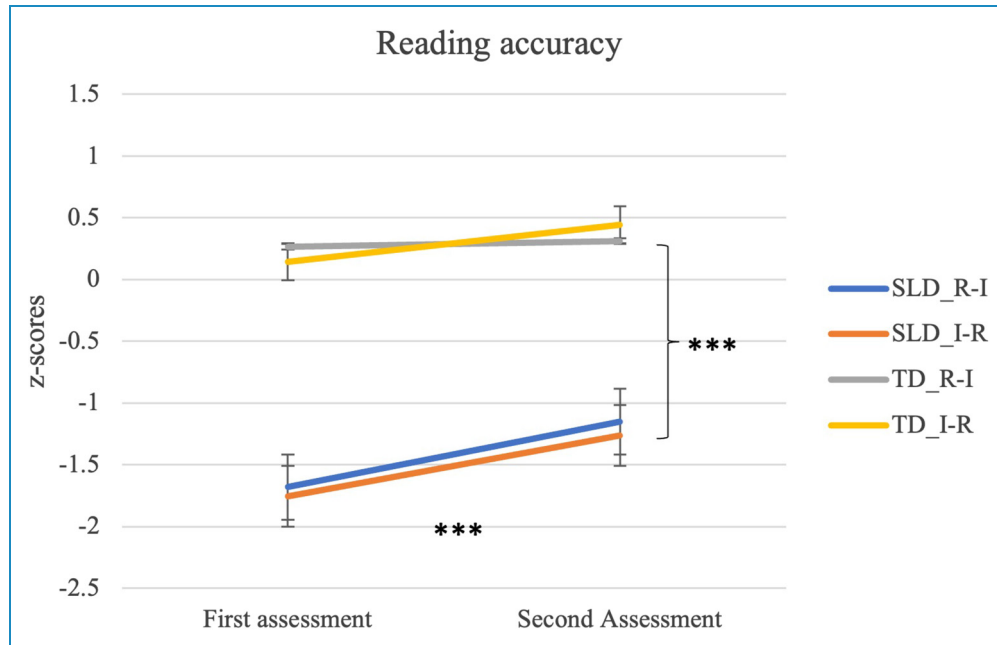
Figure 1. Repeated-measure ANOVA: Reading accuracy. SLD_I-R: children with SLD in-person vs. remote; SLD_R-I: children with SLD remote vs. in-person; TD_I-R: typically developing children in-person vs. remote; TD_R-I: typically developing children with SLD remote vs. in-person. *** $p < 0.001$.
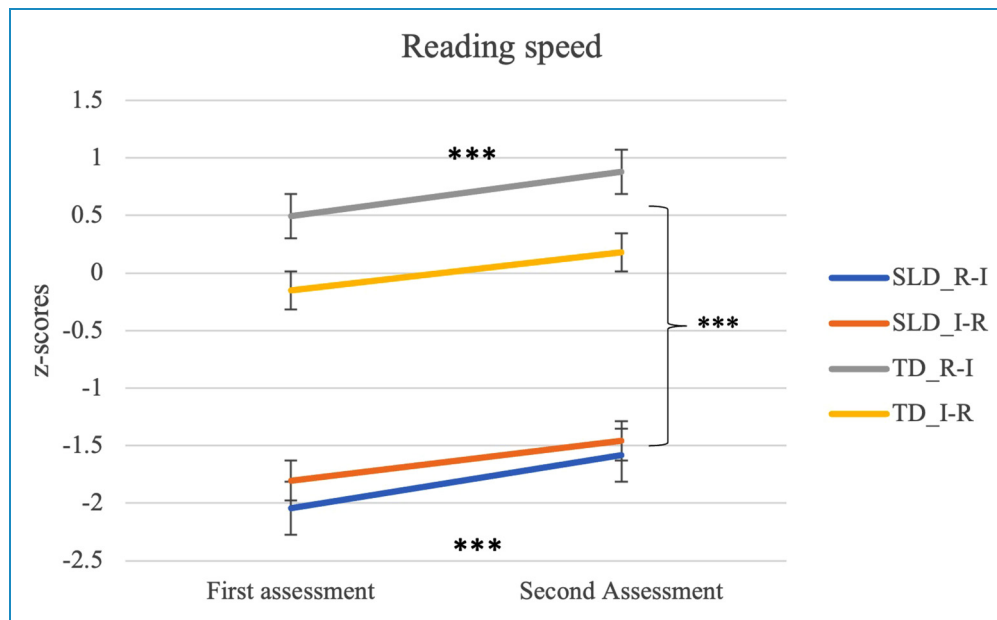

Figure 2. Repeated-measure ANOVA: Reading speed. I-R: children who underwent first in-person assessment and then remote assessment irrespectively from the diagnosis; R-I: children who underwent remote assessment first and then in-person assessment irrespectively from the diagnosis. *** $p < 0.001$.

interaction between assessment type, diagnosis, and administration order was significant ($F_{1,71} = 6.289$, $p = 0.014$, $\eta^2 = .081$), suggesting that the practice effect was modulated by the diagnosis. In particular, scores obtained by the SLD_R-I group differed according to the assessment type

(SLD_R-I: $p < 0.001$), with the in-person assessment yielding better scores than the remote home-based videoconference assessment (see Figure 4). On the contrary, scores of the TD_I-R, TD_R-I, and SLD_I-R groups did not change between the first and second assessment regardless
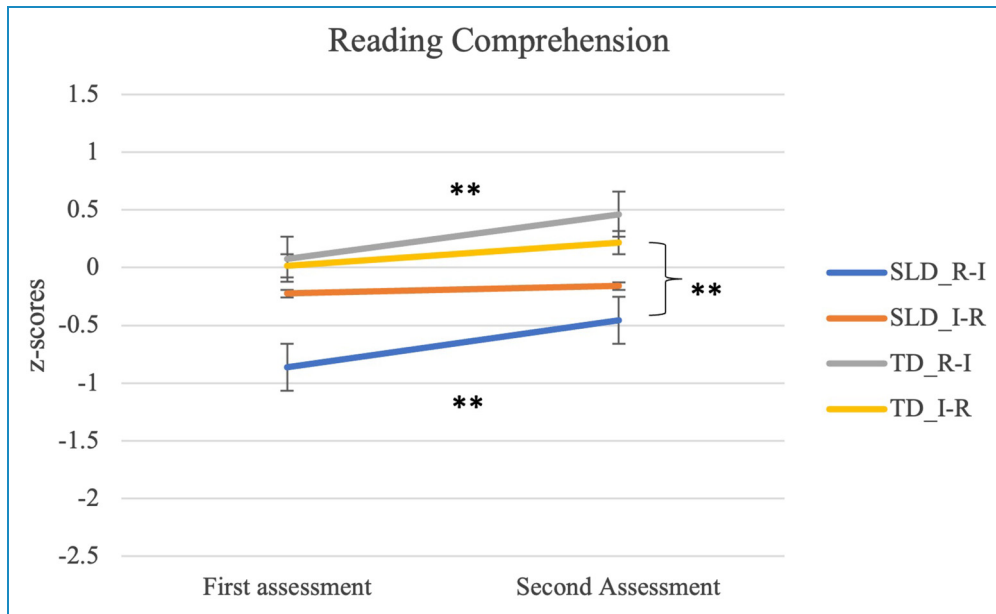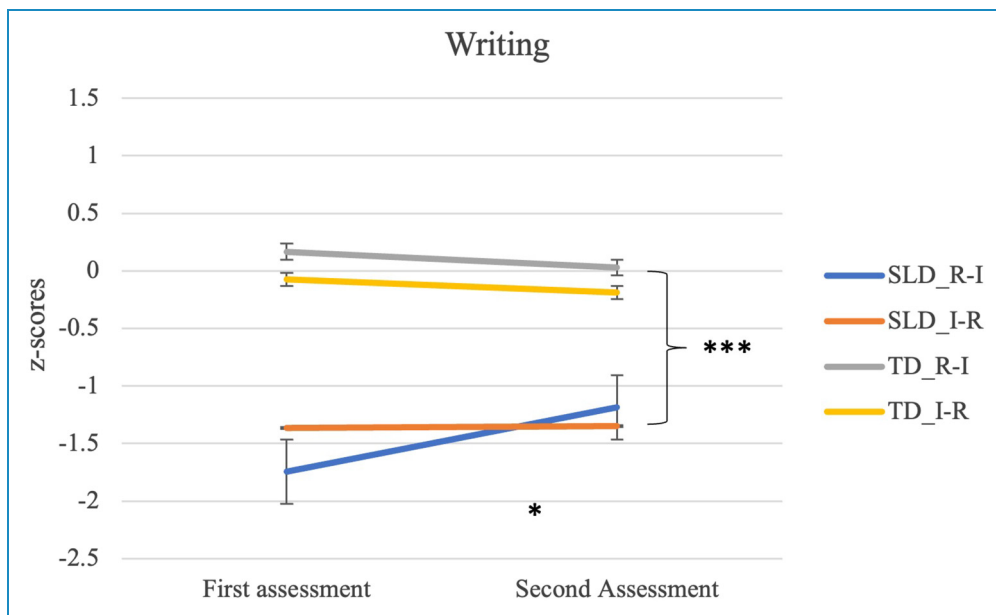
Figure 3. Repeated-measure ANOVA: Reading comprehension. I-R: children who underwent first in-person assessment and then remote assessment irrespectively from the diagnosis; R-I: children who underwent remote assessment first and then in-person assessment irrespectively from the diagnosis. ** $p < 0.01$.

Figure 4. Repeated-measure ANOVA: Writing. SLD_I-R: children with SLD in-person vs. remote; SLD_R-I: children with SLD remote vs. in-person; TD_I-R: typically developing children in-person vs. remote; TD_R-I: typically developing children with SLD remote vs. in-person. * $p < 0.05$; *** $p < 0.001$.

of the assessment type (TR_I-R: $p = 0.476$, TR_R-I: $p = 0.377$, SLD_I-R: $p = 0.928$).

In order to control for practice effect and the significant diagnosis effect, four separate between-group t-tests were used. Results showed that there were no significant differences between the two assessment types within each time of administration (children with SLD at time 1: $t_{(37)} = .451$, $p = 0.655$; children with SLD at time 2: $t_{(35)} = -.446$, $p = 0.658$; TRs at time 1: $t_{(37)} = -.958$, $p = 0.344$; TRs at time 2: $t_{(38)} = -.422$, $p = 0.676$). These

findings suggested that, within each timepoint, the two different assessment types did not lead to any difference in the subjects' scores.

Numerical processing

A main effect of diagnosis was found ($F_{1,70} = 49.300$, $p < 0.001$, $\eta^2 = .413$) suggesting that children with SLD had overall lower scores than TD (SLD: M = −.680, SE = .113; TD: M = .425, SE = .105). Moreover, the interaction between assessment type and administration order was significant ($F_{1,70} = 15.239$, $p < 0.001$, $\eta^2 = .179$). Within the two administration orders (I-R and R-I), the difference between the two assessment types was significant (I-R: $p = 0.017$ and R-I: $p = 0.003$), suggesting the presence of a practice effect (see Figure 5).

In order to control for practice effect, two separate between-group t-tests were used. Results showed that, regardless of the diagnosis, there were no significant differences between the two assessment types within each time of administration (assessment at time 1, $t_{(73)} = −.251$, $p = 0.803$ and assessment at time 2, $t_{(77)} = −.817$, $p = 0.416$). These findings suggested that, within each timepoint, the two different assessment types did not lead to any difference in the subjects' scores.

Computation

A main effect of diagnosis was found ($F_{1,70} = 54.083$, $p < 0.001$, $\eta^2 = .436$) suggesting that children with SLD had overall lower scores than TD (SLD: M = −1.393, SE = .152; TD: M = .163, SE = .141). Moreover, the interaction between assessment type and administration order showed a trend toward significance ($F_{1,70} = 3.594$, $p = 0.062$, $\eta^2 = .049$) (see Figure 6).

Semantic numerical skills

A main effect of diagnosis was found ($F_{1,69} = 31.458$, $p < 0.001$, $\eta^2 = .313$) suggesting that children with SLD had overall lower scores than TD (SLD: M = −.881, SE = .118; TD: M = .044, SE = .110). Moreover, the interaction between assessment type and administration order was significant ($F_{1,69} = 10.808$, $p = 0.002$, $\eta^2 = .135$). Within the two administration orders (I-R and R-I), the difference between the two assessment types was significant only in the R-I order ($p = 0.001$) and not in the I-R order ($p = 0.197$). These results suggest the presence of a practice effect only in the R-I order, with the second administration performed in-person yielding better scores than the first administration performed remotely (see Figure 7).

In order to control for practice effect, two separate between-group t-tests were used. Results showed that, regardless of the diagnosis, there were no significant differences between the two assessment types within each time of administration (assessment at time 1, $t_{(72)} = .278$, $p = 0.782$ and assessment at time 2, $t_{(77)} = −1.249$, $p = 0.215$). These findings suggested that, within each timepoint, the two different assessment types did not lead to any difference in the subjects' scores.
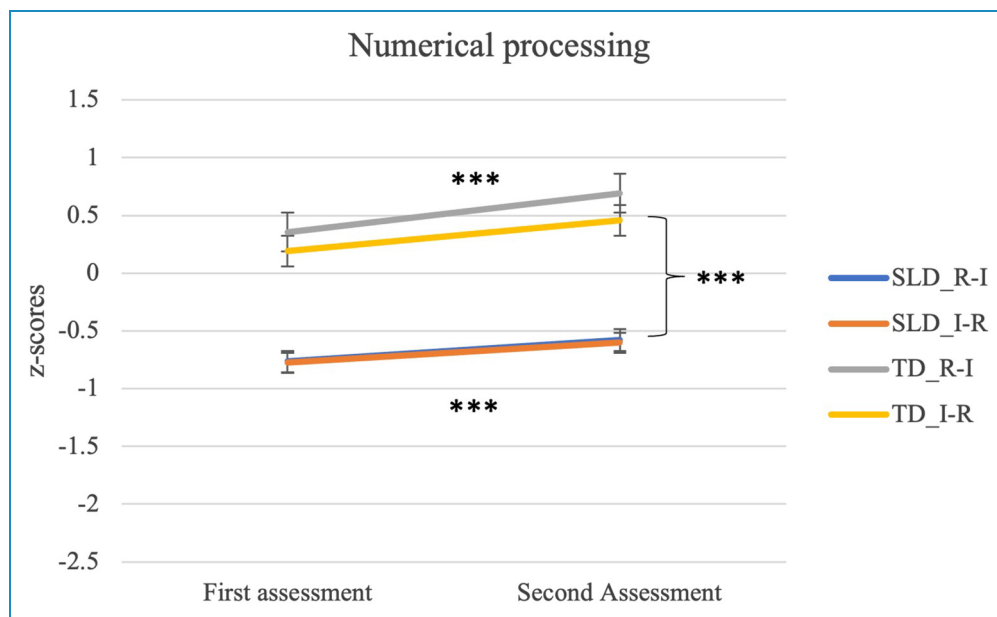

Figure 5. Repeated-measure ANOVA: Numerical processing. I-R: children who underwent first in-person assessment and then remote assessment irrespectively from the diagnosis; R-I: children who underwent remote assessment first and then in-person assessment irrespectively from the diagnosis. *** $p < 0.001$.
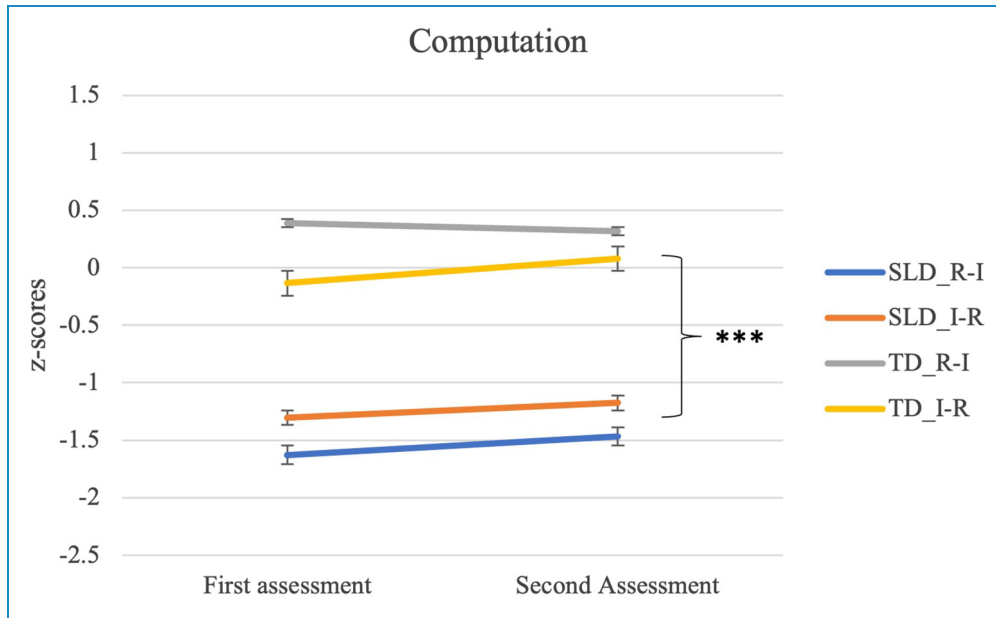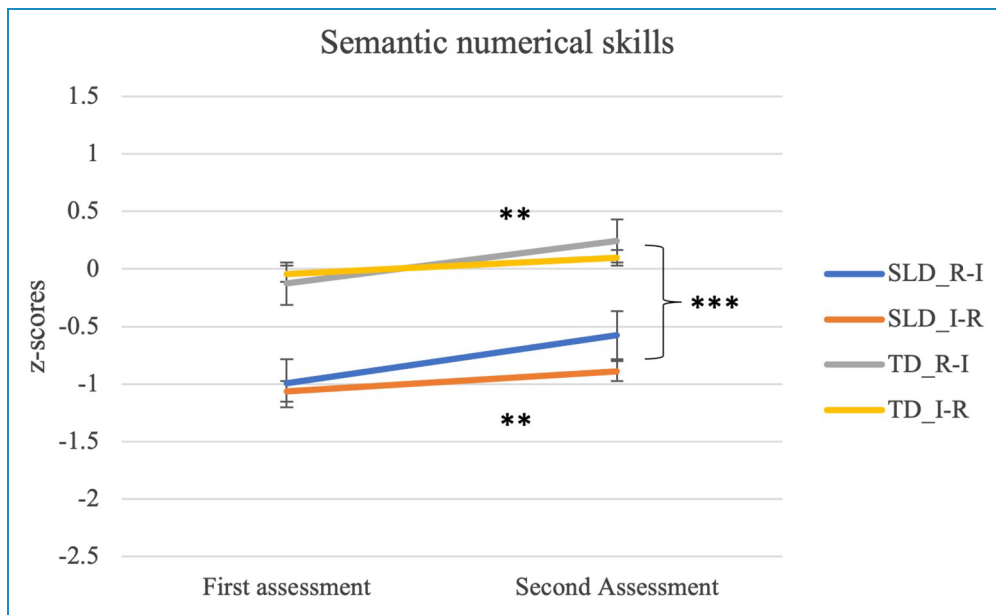
Figure 6. Repeated-measure ANOVA: Computation. I-R: children who underwent first in-person assessment and then remote assessment irrespectively from the diagnosis; R-I: children who underwent remote assessment first and then in-person assessment irrespectively from the diagnosis. *** $p < 0.001$.

Figure 7. Repeated-measure ANOVA: Semantic numerical skills. I-R: children who underwent first in-person assessment and then remote assessment irrespectively from the diagnosis; R-I: children who underwent remote assessment first and then in-person assessment irrespectively from the diagnosis. ** $p < 0.01$; *** $p < 0.001$.

Discussion

The purpose of this study was to investigate the validity and reliability of a remote home-based videoconference assessment of learning-related neuropsychological domains (i.e., reading accuracy, speed and comprehension, writing,

numerical processing, computation, and semantic numerical skills). We accordingly tested (1) whether there were significant differences between the scores obtained by each participant during two different assessment types administered during separate testing sessions which happened 2–10 days apart, and (2) whether there were differences between the

two assessment types within each time of administration (comparing R-I vs. I-R groups in the first evaluation and in the second evaluation separately). This study therefore systematically compared two assessment types (remote home-based videoconference vs. in-person) of several learning-related neuropsychological skills in both TD children and children with a clinical diagnosis of SLD.

Taken together, our findings showed that the remote home-based videoconference assessment is as reliable as vis-à-vis assessment in discriminating children with SLD from TD and could therefore represent a viable option that clinicians can adopt in their practice (see Table 2). As expected, a significant main effect of the diagnosis across all the investigated learning-related neuropsychological domains emerged, suggesting that children with SLD obtained significantly lower scores compared to TD children regardless of the assessment type and the administration order. In addition, with the I to R adaptation implemented here, both in-person and remote home-based videoconference assessment types can be considered valid and reliable for evaluating reading accuracy and speed, and numerical processing and computation. Yet, a note of caution is required regarding the validation of writing under dictation, reading comprehension, and semantic numerical sense skills. For these tasks, our results showed a potential bias against the remote home-based videoconference assessment type. This might be better explained by the changes to the administration that were deemed necessary during the remote implementation of these tests due to software limitations, and/or by technical problems (e.g., insufficient bandwidth availability).[28] Adjustments to the remote implementation of the test used to assess these skills and the adoption of more ecological technical devices to provide answers (e.g., touch screen pen) could therefore be necessary in order to make the assessment of these abilities as reliable and valid as the in-person evaluation, but would entail additional costs. However, no significant difference between the two assessment types was found within each time of administration, suggesting an overlap between the scores obtained with the two assessment types.

Moreover, we observed a significant practice effect for almost all the learning-related domains. More specifically, regardless of assessment type and consistently across groups, the scores obtained during the second assessment were always higher compared to those obtained in the first one for "numerical processing" and "reading speed." For other learning-related neuropsychological domains (i.e., reading accuracy, reading comprehension, writing, and semantic numerical sense), the practice effect was not equal among the different experimental groups. Regarding "reading accuracy," scores obtained by TD subjects did not change between first and second assessment regardless of the order followed in the administration of the different assessment types. The absence of a practice effect in TD children may reflect a ceiling effect achieved from the first

assessment. On the contrary, both SLD_I-R and SLD_R-I groups showed higher scores during the second evaluation showing a significant practice effect regardless of the order followed in the administration of the different assessment types. A similar pattern of results was obtained for "writing" assessment, where practice effect was not found in TD children, possibly reflecting a ceiling effect; however, this time only the SLD_R-I group did show a significant practice effect, which might be indicative of a potential confounding effect of the remote home-based videoconference assessment. According to previous studies reporting difficulties during remote videoconference test administration procedure for tasks involving dictation or oral repetition,[28] we can hypothesize that the observed group differences might be due to technological issues, such as the stability of the internet connection or the functionality of computer devices. Analogously, "Reading comprehension" and "semantic numerical sense' showed a significant practice effect only in the R-I groups, but not in the I-R ones, irrespectively from the diagnosis. This result suggests the presence of a potential bias against the remote home-based videoconference assessment type, for which participants' performance is hindered. The difference in administration between the in-person and the remote home-based videoconference assessments was greater in this neuropsychological domain compared to the others (i.e., each item appeared on the screen one at a time, children could not correct their responses or turn back to the previous item, answers were provided by using the mouse/touchpad; cf. "Changes in the tests' administration during the two assessement types" paragraph) and might have led to group differences in the practice effect. We can therefore hypothesize that these differences in the remote implementation of this test made the remote home-based videoconference assessment more difficult and time-consuming compared to the in-person assessment. Finally, computation did not show a significant practice effect for any group. This result could be explained by the requests of the tests included in this neuropsychological domain in which children were required to mentally solve calculations in the shortest possible time.

The current study is not without limitations. Firstly, we did not include a third experimental condition during which children underwent an in-person computerized assessment. This prevented us from clearly concluding whether our findings reflected the effect of different assessment types (i.e., remote home-based videoconference vs. in-person) or they reflected the effect of computerized vs. face-to-face with an examiner. Replication studies are therefore needed. Secondly, we did not ascertain the stability of internet connection before starting the evaluation. Previous studies showed that technological issues, such as stability of internet connection and functionality of personal devices, could influence the subjects' performance.[28] Thirdly, although we asked parents to identify a quiet room for remote home-based videoconference assessment, the setting in which this assessment took

Table 2. Repeated-measure-ANOVAs and between-groups t-tests upon the assessed neuropsychological domains.

	Do children with SLD have overall lower scores than TDs?[a]	Is there a learning effect?[b]	Are differences between assessment types modulated by the diagnosis?[c]	Is the learning effect modulated by the diagnosis?[d]	Are the scores obtained in the different assessment types overlapping within each time of administration?[e]
Reading Accuracy	44.210[112]***	17.529*** (only for children with SLD)	0.739	3.772°	✓
Reading Speed	65.620***	90.031***	0.134	0.306	✓
Reading Comprehension	8.180**	6.934** (only for R-I)	0.130	0.072	✓
Writing	16.025***	1.069	3.048	6.289* (only for the SLD_R-I group)	✓
Numerical Processing	49.300***	15.239***	0.047	0.843	✓
Computation	54.083***	3.594§	1.771	0.341	n.a.
Semantic Numerical Sense	31.458***	10.808** (only for R-I)	0.001	0.058	✓

[a]As supported by a significant main effect of the diagnosis in the repeated-measure ANOVA. F-statistic was reported.
[b]As supported by a significant interaction between assessment type and administration order in the repeated-measure ANOVA. F-statistic was reported.
[c]As supported by a significant interaction between assessment type and diagnosis in the repeated-measure ANOVA. F-statistic was reported.
[d]As supported by a significant interaction between assessment type, administration order, and diagnosis in the repeated-measure ANOVA. F-statistic was reported.
[e]As supported by a not-significant difference between the two assessment types within each time of administration in the t-tests.
*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.
° $p = 0.056$; § $p = 0.062$.
n.a.: not available.

place was not standardized. Defining the criteria of the setting where the remote home-based videoconference assessment must be carried out would be helpful in order to reduce the environmental effects on performance. Fourthly, we did not collect information regarding the acceptability and pleasantness of the remote home-based videoconference administration for children, parents, and clinicians. Such information is crucial before proposing remote home-based videoconference assessment of learning-related neuropsychological skills as a reliable tool for the clinical practices.[40–43] Fifthly, we did not randomize children with SLD. Regardless of this methodological limitation, no significant differences emerged between SLD_I-R and SLD_R-I groups when comparing the diagnosis. Sixthly, only children who have access to a compatible device could enroll in the study; hence, our results might not be generalizable to children from lower socioeconomic households. However, the devices required for the implemented remote home-based videoconference assessment are commonly used and available at an affordable cost. Finally, although the overall study population was quite small, post hoc power calculation was conducted using GPower.[44] The analysis was modeled for a repeated-measures ANOVA, a sample size of 79 subjects, four groups with two measurements, an average effect size (η^2) of the practice effect of 0.17, an average bivariate correlation between the two measurements of 0.78, and alpha equal to 0.05. Under these assumptions, the statistical power was above 90%.

Conclusion

The implementation of telemedicine's procedures during the clinical practices may offer opportunities to improve cost-effectiveness and accessibility of the health system. In the present study, we provided initial evidence about the use of remote home-based videoconference assessment for the evaluation of learning-related neuropsychological domains. While our data support the use of the remote home-based videoconference assessment as a reliable and valid tool for evaluating reading accuracy and speed, numerical processing, and computation, some cautions are required when assessing reading comprehension, writing, and semantic numerical skills. However, cumulatively, the results of this study suggest the validity and reliability of the assessment of children's learning skills via a remote home-based videoconference system. New studies aimed at systematically investigating the reliability and validity of the remote home-based videoconference assessment of learning-related neuropsychological skills must gain momentum and be of the utmost importance for clinical practices. Implementing telemedicine as an assessment tool has the potential to increase timely access to primary health care for many children.

ORCID iDs: Chiara Dondena ⓘD https://orcid.org/0000-0002-7510-2867
Sara Mascheretti ⓘD https://orcid.org/0000-0002-0060-3261

References

1. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. 5th ed. Arlington, VA: Author, 2013.
2. Sherwood AR and MacDonald B. A teleneuropsychology consultation service model for children with neurodevelopmental and acquired disorders residing in rural state regions. *Arch Clin Neuropsychol* 2020; 35: 1196–1203.
3. Luxton DD, Pruitt LD and Osenbach JE. Best practices for remote psychological assessment via telehealth technologies. *Prof Psychol Res Practice* 2014; 45: 27–35.
4. Brearly TW, Shura RD, Martindale SL, et al. Neuropsychological test administration by videoconference: a systematic review and meta-analysis. *Neuropsychol Rev* 2017; 27: 174–186.
5. Wright JH and Caudill R. Remote treatment delivery in response to the COVID-19 pandemic. *Psychother Psychosom* 2020; 89: 130–132.
6. Hodge MA, Sutherland R, Jeng K, et al. Literacy assessment via telepractice is comparable to face-to-face assessment in children with Reading difficulties living in rural Australia. *Telemed J E Health* 2019; 25: 279–287.
7. Hodge MA, Sutherland R, Jeng K, et al. Agreement between telehealth and face-to-face assessment of intellectual ability in

children with specific learning disorder. *J Telemed Telecare* 2019; 25: 431–437.

8. Ciemins EL, Holloway B, Jay Coon P, et al. Telemedicine and the mini-mental state examination: assessment from a distance. *Telemed J E Health* 2009; 15: 476–478.

9. Grosch MC, Gottlieb MC and Cullum CM. Initial practice recommendations for teleneuropsychology. *Clin Neuropsychol* 2011; 25: 1119–1133.

10. Cullum CM, Hynan LS, Grosch M, et al. Teleneuropsychology: evidence for video teleconference-based neuropsychological assessment. *J Int Neuropsychol Soc* 2014; 20: 1028–1033.

11. Galusha-Glasscock JM, Horton DK, Weiner MF, et al. Video teleconference administration of the repeatable battery for the assessment of neuropsychological status. *Arch Clin Neuropsychol* 2016; 31: 8–11.

12. Krach SK, Paskiewicz TL and Monk MM. Testing our children when the world shuts down: analyzing recommendations for adapted tele-assessment during COVID-19. *J Psychoeduc Assess* 2020; 38: 923–941.

13. Chester A and Glass CA. Online counselling: a descriptive analysis of therapy services on the internet. *Br J Guid Counc* 2006; 34: 145–160.

14. Myers K and Cain S. Practice parameter for telepsychiatry with children and adolescents. *J Am Acad Child Adolesc Psychiatry* 2008; 47: 1468–1483.

15. Backhaus A, Agha Z, Maglione ML, et al. Videoconferencing psychotherapy: a systematic review. *Psychol Ser* 2012; 9: 111–131.

16. Gros DF, Morland LA, Greene CJ, et al. Delivery of evidence-based psychotherapy via video telehealth. *J Psychopathol Behav Assess* 2013; 35: 506–521.

17. Kitaigorodsky M, Loewenstein D, Curiel Cid R, et al. A tele-neuropsychology protocol for the cognitive assessment of older adults during COVID-19. *Front Psychol* 2021; 12: 651136.

18. Farmer RL, McGill RJ, Dombrowski SC, et al. Teleassessment with children and adolescents during the coronavirus (COVID-19) pandemic and beyond: practice and policy implications. *Prof Psychol Res Pract* 2020; 51: 477–487.

19. Sperling S, Andretta CRL, Basso J, et al. Telehealth for supporting referrals to specialized care during COVID-19. *Telemed J E Health* 2022; 28: 544–550.

20. Myers CR. Using telehealth to remediate rural mental health and healthcare disparities. *Issues Ment Health Nurs* 2019; 40: 233–239.

21. Fairweather GC, Lincoln MA and Ramsden R. Speech-language pathology telehealth in rural and remote schools: the experience of school executive and therapy assistants. *Rural Remote Health* 2017; 17: 1–13.

22. Harder L, Hernandez A, Hague C, et al. Home-based pediatric teleneuropsychology: a validation study. *Arch Clin Neuropsychol* 2020; 35: 1266–1275.

23. Ruffini C, Tarchi C, Morini M, et al. Tele-assessment of cognitive functions in children: a systematic review. *Child Neuropsychol* 2022; 28: 709–745.

24. Lingely-Pottie P and McGrath PJ. A therapeutic alliance can exist without face-to-face contact. *J Telemed Telecare* 2006; 12: 396–399.

25. Jacobsen SE, Sprenger T, Andersson S, et al. Neuropsychological assessment and telemedicine: a preliminary study examining the reliability of neuropsychology services performed via telecommunication. *J Int Neuropsychol Soc* 2003; 9: 472–478.

26. Kirkwood KT, Peck DF and Bennie L. The consistency of neuropsychological assessments performed via telecommunication and face to face. *J Telemed Telecare* 2000; 6: 147–151.

27. Cowan KE, McKean AJ, Gentry MT, et al. Barriers to use of telepsychiatry: clinicians as gatekeepers. *Mayo Clin Proc* 2019; 94: 2510–2523.

28. Waite MC, Theodoros DG, Russell TG, et al. Assessment of children's literacy via an internet-based telehealth system. *Telemed J E Health* 2010; 16: 564–575.

29. Hamner T, Salorio C, Kalb L, et al. Equivalency of in-person versus remote assessment: WISC-V and KTEA-3 performance in clinically referred children and adolescents. *J Int Neuropsychol Soc* 2022; 28: 835–844.

30. Plowman L, Stephen C and McPake J. Supporting young children's learning with technology at home and in preschool. *Res Pap Educ* 2010; 25: 93–113.

31. Riva V, Mozzi A, Forni D, et al. The influence of DCDC2 risk genetic variants on reading: testing main and haplotypic effects. *Neuropsychologia* 2019; 130: 52–58.

32. Cattell RB. Intelligence: its structure, growth, and action. *Adv Psychol* 1987; 35: 14–59.

33. Wechsler D. *Wechsler Intelligence Scale for Children*. 4th ed. Firenze: Giunti Psychometrics, 2012.

34. Cornoldi C, Carretti B, Pra Baldi A, et al. *Prove MT −3 Clinica*. Firenze: Giunti Psychometric, 2020.

35. Sartori G, Job R and Tressoldi PE. *Batteria per la Valutazione della Dislessia e della Disortografia Evolutiva-2*. Firenze: Giunti Psychometrics, 2007.

36. Biancardi A, Bachmann C and Nicoletti C. *BDE 2 - Batteria discalculia evolutiva, Test per la diagnosi dei disturbi dell'elaborazione numerica e del calcolo in età evolutiva – 8–13 anni*. Trento: Erickson, 2022.

37. Peirce J, Gray JR, Simpson S, et al. Psychopy2: experiments in behavior made easy. *Behav Res Methods* 2019; 51: 195–203.

38. https://pavlovia.org/

39. IBM Corp. *IBM SPSS Statistics for Windows; Version 28.0*. Armonk, NY: IBM Corp, 2017.

40. Hajesmaeel-Gohari S and Bahaadinbeigy K. The most used questionnaires for evaluating telemedicine services. *BMC Med Inform Decis Mak* 2021; 21: 1–11.

41. Christensen LF, Gildberg FA, Sibbersen C, et al. Videoconferences and treatment of depression: satisfaction score correlated with number of sessions attended but not with age. *Telemed J E Health* 2020; 26: 898–904.

42. Morgan DG, Kosteniuk J, Stewart N, et al. The telehealth satisfaction scale: reliability, validity, and satisfaction with telehealth in a rural memory clinic population. *Telemed J E Health* 2014; 20: 997–1003.

43. Imlach F, McKinlay E, Middleton L, et al. Telehealth consultations in general practice during a pandemic lockdown: survey and interviews on patient experiences and preferences. *BMC Primary Care* 2020; 21: 1–14.

44. Erdfelder E, Faul F and Buchner A. GPOWER: a general power analysis program. *Behav Res Methods Instrum Comput* 1996; 28: 1–11.