Journal of the International Neuropsychological Society (2024), **30**, 954–965
doi:10.1017/S1355617724000432

Research Article

Tele-neuropsychology in memory clinic settings: Reliability and usability of videoconference-based neuropsychological testing

Elke Butterbrod[1,2], Dominique M.J. van den Heuvel[3,4], Pia Zevenhoven[2], Lisa Waterink[2], Mardou van Leeuwenstijn[2], Roos J. Jutten[5], Wiesje M. van der Flier[2,6] and Sietske A.M. Sikkes[1,2]

[1]Department of Clinical, Neuro- and Developmental Psychology, Faculty of Movement and Behavioral Sciences, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands, [2]Alzheimer Center Amsterdam, Department of Neurology, Amsterdam University Medical Center, Amsterdam, the Netherlands, [3]Department of Neurology, Basalt Rehabilitation Center, Leiden, The Netherlands, [4]Department of Health, Medical and Neuropsychology, Faculty of Social and Behavioral Sciences, Leiden University, Leiden, The Netherlands, [5]Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA and [6]Department of Epidemiology and Data Science, Vrije Universiteit Amsterdam, Amsterdam UMC location VUmc, Amsterdam, the Netherlands

Abstract

Objective : Neuropsychological assessment through VideoTeleConferencing (VTC) can help improve access to diagnostic and follow-up care in memory clinics. This study investigated the stability of performance on VTC assessment in relation to in-person assessment using a test-retest design and explored user experiences of VTC assessment. **Materials and Methods**: Thirty-one patients (62 ± 6.7 years, 45% female, 58% Subjective Cognitive Decline, 42% Mild Cognitive Impairment/dementia diagnosis) were included from the Amsterdam Dementia Cohort between August 2020 and February 2021. Patients underwent a face-to-face neuropsychological assessment followed by a VTC assessment using the same test protocol within 4 months. Reliability coefficients were calculated using intraclass correlation coefficients (ICC). For each test, the proportion of clinically relevant differences in performances between assessment modalities was calculated. User experiences of patients and neuropsychologists were assessed with questionnaires (User Satisfaction and Ease of use [USE] questionnaire and System Usability Scale [SUS]). Neuropsychologists also participated in a focus group. **Results:** ICC values were moderate to excellent (0.63-0.93) for all test measures in the total sample. On all tests, most patients did not show clinically relevant performance differences between modalities. Patients and neuropsychologists reported overall positive VTC system usability, although neuropsychologists indicated in the focus group that patients without cognitive impairment required less training for the system and were more independent. **Conclusion**: VTC assessment showed adequate to excellent test-retest reliability for a broad range of neuropsychological tests commonly used in practice. Assessment through VTC may be a user friendly method in the memory clinic, especially to monitor individuals at risk for future cognitive decline.

Keywords: Tele-neuropsychology; remote neuropsychological assessment; memory clinic; dementia; mild cognitive impairment; subjective cognitive decline

(Received 25 September 2023; final revision 31 July 2024; accepted 11 August 2024; First Published online 21 October 2024)

Introduction

The reduction in opportunities for in-person cognitive assessments during the COVID-19 pandemic has accelerated research efforts into feasibility and reliability of remote cognitive testing, including assessment through video teleconference (VTC) (Bilder et al., 2020; Geddes et al., 2020; Marra et al., 2020; Owens et al., 2020; Pulsifer et al., 2021; Rochette, Rahman-Filipak, Spencer, Marshall & Stelmokas, 2021; Singh & Germine, 2021; Underwood et al., 2021). In addition to maintaining sufficient access to crucial neuropsychological services under exceptional circumstances (Owens et al., 2020; Underwood et al., 2021), implementing solid protocols for remote, at-home cognitive testing can also facilitate diagnostic care and monitoring in the memory clinic on the long-term by taking away several barriers to undergo (repeated) testing (Hewitt et al., 2022; Tsiakiri et al., 2022). Among supervised remote neuropsychological assessments, VTC may, for example, be preferable to telephone-based assessment for a variety of reasons, including reliability (Hunter et al., 2021). Undergoing assessment through VTC from home can be an alternative to assessment at the clinic for patients who live in rural communities (Vestal et al., 2006), struggle with health or mobility limitations, have lower motivation for testing (Castanho et al., 2014) or experience anxiety surrounding clinic visits (that may also affect test performance (Dorenkamp et al., 2021)).

In a memory clinic setting, it is crucial to know which tests can be dependably administered through VTC, but also which patients

Corresponding author: Elke Butterbrod; Email: E.Butterbrod@vu.nl

Cite this article: Butterbrod E., van den Heuvel D.M.J, Zevenhoven P., Waterink L., van Leeuwenstijn M., Jutten R.J., van der Flier W.M., & Sikkes S.A.M. (2024) Tele-neuropsychology in memory clinic settings: Reliability and usability of videoconference-based neuropsychological testing. *Journal of the International Neuropsychological Society*, **30**: 954–965, https://doi.org/10.1017/S1355617724000432

are more (or less) suitable for and open to undergoing this method of assessment. Research has shown that VTC assessment is likely feasible in community-dwelling older individuals (Hildebrand et al., 2004), individuals with self-perceived cognitive decline but no major impairment (Gnassounou et al., 2021), and even those with cognitive impairment (Cullum et al., 2014; Hunter et al., 2021; Parikh et al., 2013; Wadsworth et al., 2018). At the same time, the reliability and validity of neuropsychological testing administered through VTC in a clinic-to-home setting are not yet adequately established (Marra et al., 2020; Parsons et al., 2022), in particular for complete, comprehensive assessment protocols used for regular clinical diagnostics. Furthermore, we do not have an adequate picture of other important issues relating to the integration VTC in memory clinic neuropsychological services for individuals with varying levels of cognitive impairment (Bilder et al., 2020; Hildebrand et al., 2004), specifically user experiences of both administrators and patients.

Research investigating reliability of VTC assessment commonly focuses on tests that have existing adaptations for remote administration (König et al., 2021) or are based on verbal instruction and response, such as verbal memory and fluency tests. Such tests are relatively easy to implement in remote assessment and appear to have largely sufficient reliability metrics (Brearly et al., 2017; Cullum et al., 2014; Fox-Fuller et al., 2022) and produce scores that are not significantly different from face-to-face assessment (Marra et al., 2020), although test-retest reliability of remote assessment may deviate slightly from in-person administration (Fox-Fuller, Ngo, et al., 2022; Marra et al., 2020). VTC administration of other tests that are often part of a neuropsychological workup, such as timed tests, tests relying on visual or motor modalities or those that tap into executive functions may come with challenges concerning administration, accuracy and performance monitoring (Bloch et al., 2021; Fox-Fuller, Ngo, et al., 2022). Findings regarding reliability of such tests and the deviation in scores compared to regular administration vary, are sometimes based on adapted versions of the original or too limited for clinical implementation (Brearly et al., 2017; Hunter et al., 2021; Lunardini et al., 2020; Marra et al., 2020; Owens et al., 2020; Wadsworth et al., 2018).

Alongside sufficient reliability, an important part in evaluating the value of VTC assessment involves the user experiences from both administrators and patients (Appleman et al., 2021; Parsons et al., 2022). A survey among US-based neuropsychological test administrators (Fox-Fuller et al., 2022), identified connection issues, limited technological access and external distractions as some of the major challenges. Such challenges are likely to exist across countries and settings (Hewitt et al., 2022; Sumpter et al., 2023). Still, in memory clinics we may also encounter challenges inherent to the patient population, including high average age, varying technological literacy, and substantial cognitive and functional decline. All of these factors potentially affect patients' ability to navigate VTC systems as well as the cognitive profile derived from remote assessment (Brearly et al., 2017; Parks et al., 2021). More research is needed to understand how previously identified challenges with VTC relate to the memory clinic context, and to gauge specific challenges and opportunities at play.

The aims of the current study were to 1) investigate the stability of performances on a VTC assessment comprised of commonly used, non-adapted neuropsychological tests in relation to in-person assessment using a test-retests design, and 2) evaluate user experiences of neuropsychologists and cognitively impaired and unimpaired patients of a Dutch memory clinic.

Methods

Participants and study design

The current study is a mixed-methods observational prospective study, with a test-retest design. Participants considered for inclusion were older adults who visited Alzheimer Center Amsterdam between August 2020 and February 2021 for a multidisciplinary diagnostic evaluation. The current diagnostic workup includes a neurological evaluation, neuroimaging, lumbar puncture and face-to-face in-clinic neuropsychological assessment. Details regarding the complete diagnostic procedure and criteria are described in (van der Flier & Scheltens, 2018). The neuropsychological test protocol is described in Table 1.

Inclusion criteria were a (completed) face-to-face in-clinic neuropsychological assessment and access to a computer for VTC at home. Exclusion criteria were: insufficient knowledge of the Dutch language, and insufficient clinical condition as judged by the neuropsychologist based on information gathered at the diagnostic evaluation (e.g., advanced dementia or behavioral problems). Diagnosis was known at time of inclusion but was not a deciding factor for in- or exclusion by itself in the current study.

VTC assessments took place within 4 months after the face-to-face assessment. Data were collected between August 2020 and April 2021. The study was conducted in accordance with the Declaration of Helsinki (World Medical Association, 2013) and was approved by the Institutional Review Board of VU University Medical Center (VUmc). All participants signed informed consent.

Measures of the neuropsychological assessment

The same standardized neuropsychological assessment protocol was performed at both the in-clinic assessment and VTC assessment, except the Mini Mental State Examination (MMSE, in-clinic only). The assessment comprised a standard clinical workup used for diagnostic purposes at the Alzheimer center Amsterdam, also see van der Flier & Scheltens (2018). The protocol consisted of 19 tests, covering six cognitive domains: global cognition, memory, attention, executive function, language, and visuo-spatial function. See Table 1 for an overview of the tests as well as the measures derived from the tests that were subsequently used in the analyses.

We used both raw scores and standardized T-scores, depending on the analysis performed (see Statistical Analyses). Once standardized, the T-scores were also classified according to published guidelines for Dutch clinical practice that were adapted from the American Academy of Clinical Neuropsychology (Hendriks & Kessels, 2020). into the following performance classes: *very low/impaired* (T score <30), below average (30 ≤ T score ≤36), *low-average* (37 ≤ T score ≤ 42), *average* (43 ≤ T score ≤ 56), *high average* (57 ≤ T score ≤ 63), *above average* (64 ≤ T score ≤ 71), and *high* (T score ≥ 72). As Dutch norms for the Letter Digit Substitution Test (LDST) (van der Elst et al., 2006) were based on the 60-seconds version of the test (as opposed to the 90-second version in our protocol) we adapted the raw scores of the LDST (0.67* score at 90 s) before standardizing. Performance classes for the Visual Association Test (VAT) and Number Location tests included *impaired* vs *unimpaired* only, due to the skewed distribution of scores (ceiling effect).

VTC assessment procedure

VTC assessment was clinic-to-home, i.e., the neuropsychologist resided at the clinic and the patient at home. For the VTC

https://doi.org/10.1017/S1355617724000432 Published online by Cambridge University Press

Table 1. Neuropsychological assessment overview with cognitive domains

Cognitive domain	Tests	Measures used
Global cognitive functioning	Mini Mental State Examination (MMSE)†	Total score
	Montreal Cognitive Assessment (MoCa)	Total score
Memory	Visual Association Task (VAT) A and B*	Total trial score A* and B*
	Rey Auditory Verbal Learning Test (RAVLT)	Direct recall*
		Delayed recall*
		False negatives
		False positives
	Rey Complex Figure (Rey CFT)	Total score on recall subtest*
Attention	Digit span forward	Total score*
	Trail Making Test A (TMT A)	Total time to complete*
	Stroop color word test I	Total time to complete*
	Stroop color word test II	Total time to complete*
Executive function	Digit span backwards*	Total score*
	Trail Making Test B (TMT B)	Total time to complete*
	Stroop color-word test (Stroop III)	Total time to complete*
	Phonemic fluency test; version D-A-T	Stroop III/II index (interference score)*
		Total correct words*
Language	VAT (naming)	Total score*
	Category fluency test; animals	Total correct words*
Visuo-spatial function	Visual Object and Space Perception (VOSP)	Total score
	-Dot Counting (remote only)	Total score
	-Fragmented letters	Total score*
	-Number location	Total score on copy subtest*
	Rey CFT	Total score on copying subtest

*This test is also used for clinical classification of performance on the corresponding domain (see statistical analyses- Clinically relevant changes in performance).
†Baseline visit only.

assessment, instructions and response forms where required (TMT, Rey Complex Figure test, LDST, MoCa) were sent to patients' homes with explicit instructions to not open these until the start of the VTC visit. Patients were instructed to use a laptop or desktop computer for VTC, and to limit risk of distractions by undergoing the assessment in a quiet room in their home and instructing cohabitants to not disturb. For patients who were unsure of how the system should be setup, a partner or family member was allowed in the room to help. Audio and video settings were tested with the neuropsychologist in a brief preparatory meeting and at the start of the neuropsychological assessment.

The VTC assessment took place via secure conferencing using Microsoft Teams®, which is the primary conferencing method at the Amsterdam UMC. For some paper-and-pencil tests that required performance monitoring, including the TMT and Rey Complex Figure test, the patient was asked to adjust their camera to allow the neuropsychologist to view their actions during the task. Technological or logistic issues and consequences for the assessment (e.g., missing test data) were noted by the neuropsychologist.

Missing data

We inspected whether one of the modalities produced more missing test scores than the other, and for what reason (e.g., if there were systematically more missing test scores due to lack of test comprehension in one modality). Neuropsychologists recorded the number and reason for missing test measures at baseline and VTC. Furthermore, they recorded any issues that arose during the VTC assessment that could be of influence on the test outcome, e.g., lagging connection. Reasons for missing data were dichotomized into 0) patient was unable to perform the test (due to lack of understanding of the instructions) or 1) not administered (e.g., due to time restrictions of the assessment). We evaluated whether the number and main reason for missing test scores were different for the assessment modalities with chi-square tests.

User friendliness and system usability

After the VTC assessment, both the participant and test administrator completed a Dutch version of the User Satisfaction and Ease of use (USE) questionnaire (Lund, 2001) and the System Usability Scale (SUS) (Brook, 1995). The SUS consists of 10 statements aimed at evaluating general usability, using a 5-point likert scale ranging from strongly disagree (score of 1) to strongly agree (score of 5). The USE questionnaire consists of 30 statements using a 7-point Likert scale, and covers domains of usefulness, ease of use, ease of learning and satisfaction. A focus group was held with the neuropsychologists (LW, RJ, PZ) after the final assessment for an evaluation of their experiences with the VTC assessment, including main issues, opportunities and the experience with testing patients with different diagnoses.

Statistical analyses

Statistical analyses were performed with SPSS, version 27. A p-value of .05 was considered statistically significant, unless otherwise specified.

Descriptive statistics (mean, median, proportions, depending on the variable) were used to describe sociodemographic and clinical characteristics. Comparisons between diagnostic groups (Subjective Cognitive Decline [SCD] versus Mild Cognitive Impairment [MCI]/dementia) were made using independent samples-tests and chi-square tests, as appropriate.

Stability of performances across modalities

Raw scores on the test measures were used for investigate performance stability.

Intraclass Correlation Coefficients (ICC) were computed (type absolute agreement, two-way-mixed model) to measure homogeneity of scores across baseline vs. the VTC assessment for each test with continuous scores. We adopted published cutoffs to evaluate the ICC values as "poor" (<0.50), "moderate" (0.50–0.74),

https://doi.org/10.1017/S1355617724000432 Published online by Cambridge University Press

"good" (0.75–0.90) or "excellent" (0.90) (Koo & Li, 2016). Tests with restricted score ranges (i.e., VAT – naming, Visual Object and Space Perception [VOSP] – Fragmented letters and Number location subtests) were considered non-continuous, in which case we calculated the percentage absolute agreement between measurements. ICC values were calculated for the total sample and plotted as a function of clinical diagnosis (SCD and MCI/dementia).

To investigate stability of rank ordering between assessments, and compare with published literature, we ran correlational analyses between the in-person baseline and VTC assessment scores (Pearson product-moment or Spearman rho, for continuous scores). For non-continuous measures, chi-square or Fisher's exact tests were performed.

We investigated differences in mean test scores from the in-person baseline to VTC assessment with paired-samples t-tests or Wilcoxon signed-rank tests, depending on data distribution. Effect sizes of the differences were calculated as Cohen's D (M difference / SD difference) or r (Wilcoxon Z / $\sqrt{}$test sample size).

For tests with continuous scores, we used Bland–Altman plots to visualize the difference in scores between both measurements against the average of both measurements for each individual. These plots indicate whether there are systematic differences between scores on the face-to-face vs the remote assessment (Altman & Bland, 1983). T-scores of the measures (see below) were used for scalability. 95% CI limits of agreement were set.

Clinical relevance of differences in performance across modalities

Standardized scores were used for the evaluation of clinical relevance of differences between modalities. For tests with continuous scores, a clinically relevant difference was defined as a difference of at least two of the previously mentioned performance classes between the measurements, e.g., below average at baseline assessment vs. average at VTC assessment. For tests with non-continuous scores, a shift between the performance classes (impaired/unimpaired) was seen as clinically relevant. We calculated the proportion of patients showing a clinically relevant difference in performances (worse or better) for each test.

Usability

Total scores on the USE questionnaire and domain scores for patients and neuropsychologists were computed and compared using Mann–Whitney U tests due to non-normal distributions. We also computed scores domain scores between the group without vs with cognitive impairment (i.e., SCD vs MCI/dementia diagnosis) as reported by patients and neuropsychologists separately. Responses on each item of the SUS were deemed categorical in nature, and statistically compared, again between the group without vs with cognitive impairment, with Chi-square or Fisher's exact tests for patients and neuropsychologists separately. Focus group responses were summarized (author: SS) and described in addition to the quantitative results.

Results

Sample

133 individuals who underwent a face-to-face assessment as part of clinical care were considered for inclusion. Thirty-two patients met the inclusion criteria and wanted to participate. Of the 101 patients who were excluded, 12 did not have the appropriate setup for VTC at home and 2 mentioned they did not think they possessed sufficient technological skills for VTC. Other main reasons for exclusion were poor clinical status or psychological symptoms (e.g., stress after diagnosis or depressive symptoms), visual or language impairments, not wanting to participate or logistical reasons (e.g., not being able to plan a VTC assessment in the appropriate timeframe).

One patient withdrew after informed consent, but before the VTC assessment. The final sample thus consisted of 31 patients (mean age 62.2 ± 6.7 years, 45.2% female, 84.2% higher education; see Table 2 for an overview of sample characteristics). The group without cognitive impairment was younger (59.7 ± 5.2 years) than the group with cognitive impairment (65.9 ± 7.0 years, $p < .01$). As expected, MMSE score was higher in the former (28.7 ± 1.3) compared to the latter group (25.5 ± 3.2, $p < .01$). There were no differences between groups with regard to other baseline characteristics. One assessment was conducted on IPad, all other assessments were conducted via laptop or personal computer.

Analysis of missing data for face-to-face and VTC administration

Supplementary table 1 provides an overview of missing data per test (in total and stratified according to primary reason for absence).

Face-to-face

Twelve patients had at least one missing test value. Four of these patients (diagnoses MCI $N = 1$; AD $N = 2$; vascular dementia $N = 1$) had not completed at least one task because of insufficient understanding or execution of the test instructions. In the eight other cases, missing data were not related to test understanding or execution, but due to time restrictions.

VTC

Nine patients had at least one missting test value. Four of these patients (diagnoses AD $N = 3$; vascular dementia $N = 1$) had not completed at least one task because of insufficient understanding or execution of the test instructions. In the other five cases, missings were not related to test understanding or execution, but due to time restrictions.

Chi-square tests indicated no difference in the total number of missing scores between the two assessments, p's < .05.

Neuropsychological assessment-stability of performances

Test scores at face to face and remote assessment and the stability indicators (change in mean scores with effect sizes, ICC, correlations) in the total sample are displayed in Tables 3 and 4. ICC values are also depicted in the total sample and stratified according to diagnosis (SCD vs MCI/dementia) in Figure 1.

All tests showed moderate to excellent ICC values (range 0.63–0.93) in the total sample. Absolute agreement (non-continuous test measures) ranged from 19% (VOSP number location) to 87% (VAT naming). Pearson and Spearman correlation indicated significant relationships (all p's < .05) between face-to-face and VTC administration for all test measures. Chi-square tests indicated similar score distributions between the VAT Naming and VOSP Fragmented Letters (p's > .05), but not for VOSP Number Location ($p < .01$).

Independent samples t-tests/Wilcoxon Signed-rank tests showed higher mean scores for the Rey Auditory Verbal Learning Test (RAVLT) recognition – false negatives (i.e., more false negatives on average), and Rey Complex Figure Test (Rey CFT) – copy condition (i.e., more correctly drawn parts of the figure) at the VTC assessment, showing small to medium effect sizes. After stratifying for diagnosis, the significant difference in the Rey CFT mean scores in the total sample was explained by a higher score among the

https://doi.org/10.1017/S1355617724000432 Published online by Cambridge University Press

Table 2. Sample characteristics

	Total ($N = 31$)	Not cognitively impaired ($N = 18$)	Cognitively impaired ($N = 13$)
Age at time of baseline (M ± SD)	62.2 ± 6.7	59.7 ± 5.2**	65.9 ± 7.0**
Male (*N*, %)	17, 54.8%	8, 44.4%	9, 69.2%
Educational level (Verhage coding, *n*,%)			
Lower	1, 3.2%	0, 0%	1, 8%
Middle	7, 22.6%	4, 22.2%	3, 23.1%
Higher	23, 84.2%	78.8%	69.2%
Diagnosis (*n*,%)			
Subjective cognitive decline (SCD)	18, 54.8%	18, 100%	n/a
Mild Cognitive Impairment (MCI)	7, 22.6%	n/a	7, 53.8%
Dementia – Alzheimer's Disease (AD)†	5, 16.1%	n/a	5, 38.4%
Dementia – Vascular	1, 3.2%	n/a	1, 7.6%
Days between assessments (M ± SD, range)	76.5 ± 22.2, 22-112	74.1 ± 22.0	79.9 ± 23.0
MMSE score at first visit (M ± SD)	27.4 ± 2.8	28.7 ± 1.3**	25.5 ± 3.2**

Note: MMSE = Mini Mental State Examination.
**Significant difference between diagnostic groups, $p < .01$.
Verhage coding can be translated as follows: lower education = elementary school up to lower vocational training (generally 6 to 13 years total education); middle education = intermediate vocational training (e.g., generally 10–14 years total education); higher education = preparatory scientific education, a bachelor's degree or higher (generally 12 to 16+ years
†With or without vascular pathology.

Table 3. Scores at the face-to-face and remote assessment

		Face-to-face			VTC			Difference	
Domain	Test	N	M	SD	N	M	SD	Mdiff‡ (pairs)	ES†
Global	MoCa	31	25.06	3.99	31	25.77	3.65	−0.71	−0.30
Memory	VAT A	31	15.52	4.73	30	15.00	4.48	0.47	−0.15
	VAT B	28	15.61	2.39	27	15.78	3.90	−0.42	−0.21
	RAVLT – direct recall	31	37.68	13.76	30	35.83	15.74	1.17	0.16
	RAVLT – delayed recall	31	6.84	4.51	30	5.77	4.97	0.90	0.30
	RAVLT – false positives	30	1.17	2.10	30	1.20	2.17	−0.03	−0.01
	RAVLT – false negatives	30	1.27	2.05	30	2.53	3.25	−0.83*	−0.43
	Rey CFT - recall	28	15.71	7.31	28	18.09	7.52	−2.25	−0.71
Attention/ speed	Digit span forward	31	12.48	2.79	30	13.23	2.66	−0.90	−0.54
	TMT part A	31	46.03	54.62	29	39.34	16.73	−2.07	−0.30
	Stroop color word test I	31	46.16	12.61	30	52.13	21.28	−5.47	−0.47
	Stroop color word test II– total time	31	68.74	25.48	30	69.00	23.30	0.63	−0.18
	Letter-Digit Symbol Substitution	27	43.74	13.64	28	43.36	11.88	1.20	0.23
Executive function	Digit span backwards	31	8.29	1.97	30	9.57	3.01	−1.33	−0.67
	TMT part B	28	100.71	65.83	27	102.37	48.38	0.59	0.13
	Stroop color word test III – total time	29	110.55	45.69	28	116.64	55.88	−4.64	−0.09
	Stroop color word test III – interference index	29	46.10	29.63	28	51.36	39.56	−4.61	−0.01
	Letter Fluency	28	38.82	12.64	29	39.45	14.64	−1.27	−0.14
Language	VAT - naming	31	11.81	.91	30	11.97	.18	−0.17	−0.15
	Category Fluency; animals	31	20.29	7.50	30	21.67	9.58	−1.83	−0.34
Visuo-spatial function	VOSP- Fragmented letters	27	18.93	1.24	29	18.69	1.83	0.16	−0.07
	VOSP- Number location	27	9.11	1.19	30	9.97	2.67	−1.00	−0.33
	Rey CFT – copy	30	33.00	4.83	29	34.59	3.24	−1.69*	−0.44

Note: BL = baseline assessment in-person, VTC = video-teleconference assessment, VAT = Visual Association Test, RAVLT = Rey Auditory Verbal Learning Test, Rey CFT = Rey Complex Figure Test, TMT = Trail Making Test, VOSP = Visual Object & Space Perception test. $P < .05$* or $p < .01$**.
‡Difference between baseline-remote assessment (pairs).
†Cohen's D (Mdiff/SDdiff) or r (Wilcoxon Z/$\sqrt{}$(test sample size)).

cognitively unimpaired patients (*SCD group* face-to-face: 34,35 ± 1.50; VTC: 35,41 ± 1.12; $p = .02$, vs. *MCI/dementia group* face-to-face: 30,8 ± 37,04; VTC: 33,42 ± 4,72; $p > .05$). The difference in RAVLT recognition - false negative scores in the total sample was significant in neither the SCD group alone (face-to-face: 0.59 ± 0.71; VTC: 1.24 ± 1.2.02; $p = .13$) nor the MCI/dementia group alone (2.33 ± 2.87; VTC: 3.42 ± 2.54, $p = .07$).

Bland–Altman plots

Differences between the assessment modalities were close to $T = 0$ for all tests except for the Rey CFT-delayed condition (mean difference $T = -10$; 1 standard deviation). We note that differences in scores between the assessments appeared fairly stable as a function of patients' mean performance across measurements for most tests (i.e., the differences in scores between face-to-face and VTC assessment did not become more larger as patients' mean performances over the assessments were lower or higher). Especially the RAVLT conditions, Stroop test 2 and Digit Span forward (verbal tests of memory, processing speed and attention), but also LDST, (a written test of visuomotor speed) showed stability. For the Rey CFT delayed and TMTa, there appeared to be less variability in difference scores when mean performance over assessments was worse. For Letter fluency the opposite seemed to be the case; variability decreased somewhat as the mean performance was better.

Bland–Altman plots for the LDST, Rey CFT delayed and Letter fluency are shown in Figure 2, and those for the other tests are shown in Supplementary figure 1.

https://doi.org/10.1017/S1355617724000432 Published online by Cambridge University Press

Table 4. Stability coefficients from baseline to VTC assessment

Domain	Test	Stability		Reliability coefficient face-to-face†
		⌝/X^2	ICC/% agreement	
Global	MoCa	0.67**	0.88	0.82
Memory	VAT A	17.15*	0.87	0.74–0.81
	VAT B	47.8**	0.66	
	RAVLT – direct recall	0.89**	0.94	0.80–0.83
	RAVLT – delayed recall	0.78**	0.88	
	RAVLT – false positives	0.53**	0.81	
	RAVLT – false negatives	0.48*	0.80	
	Rey CFT – recall	0.65**	0.71	0.59 (*r*)
Attention/ speed	Digit span forward	0.81**	0.87	0.82
	TMT part A	0.61**	0.71	0.79
	Stroop color word test I	0.74**	0.62	0.73
	Stroop color word test II – total time	0.84**	0.88	0.78
	Letter-Digit Symbol Substitution	0.91**	0.95	0.85 (*r*)
Executive function	Digit span backwards	0.69**	0.77	0.82
	TMT part B	0.70**	0.81	0.89
	Stroop color word test III – total time	0.60**	0.74	0.85
	Stroop color word test III – interference index	0.57**	0.67	–
	Letter Fluency	0.75**	0.87	0.78–0.85
Language	VAT - naming	0.07	87%	–
	Category Fluency; animals	0.83**	0.87	0.83 (*r*)
Visuo-spatial function	VOSP- Fragmented letters	12.63	19%	–
	VOSP- Number location	30.78*	36%	–
	Rey CFT – copy	0.40**	0.63	0.18 (*r*)

Note: BL = baseline assessment in-person, VTC = video-teleconference assessment, VAT = Visual Association Test, RAVLT = Rey Auditory Verbal Learning Test, Rey CFT = Rey Complex Figure Test, TMT = Trail Making Test, VOSP = Visual Object & Space Perception test.
⌝Pearson product-moment or Spearman rho.
†Reliability coefficients (ICC or r) across multiple face-to-face assessments reported in literature (Bruijnen et al., 2020; Jutten et al., 2018; Lezak et al., 2012; Schmand et al., 2008).

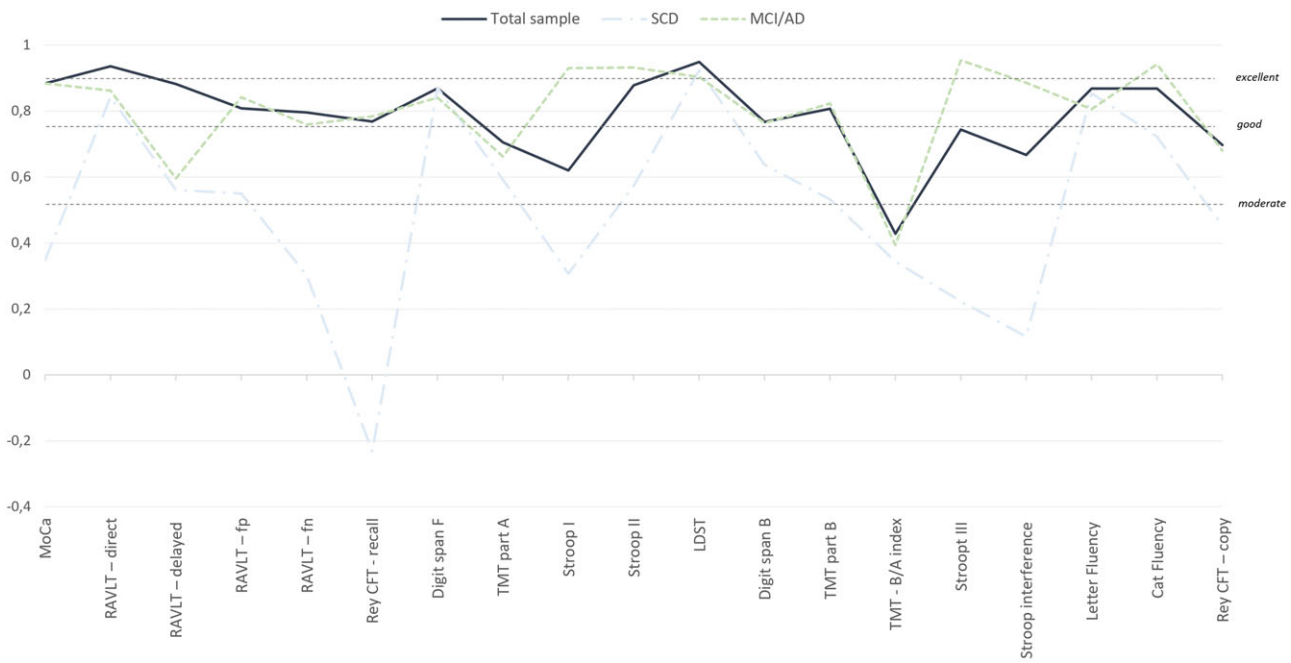

Figure 1. ICC values of the test measures with continuous scores in the total sample (dark blue), and stratified by diagnosis (Subjective Cognitive Decline [SDC] in light blue, MCI/Dementia in light green). Dotted lines indicate excellent, good and moderate ICC values. Abbrevations: VAT = Visual Association Test, RAVLT = Rey Auditory Verbal Learning Test, Rey CFT = Rey Complex Figure Test, TMT = Trail Making Test, ICC = intraclass correlation coefficient, MCI = Mild Cognitive Impairment.

Clinical interpretation of stability and differences in test scores

Supplementary figure 2 displays the amount of stability and clinically relevant shifts in performance classes from baseline to VTC follow-up per test in the sample according to the predefined cut offs.

For all tests, most patients showed no clinically relevant difference. Number Location (100%), LDST (96%) and Letter Fluency and Stroop 2 (both 93%) measures showed the highest number of stable classifications from baseline to VTC assessment. TMT A (17%) and TMT B (15%) and (their derived) were the tests

https://doi.org/10.1017/S1355617724000432 Published online by Cambridge University Press

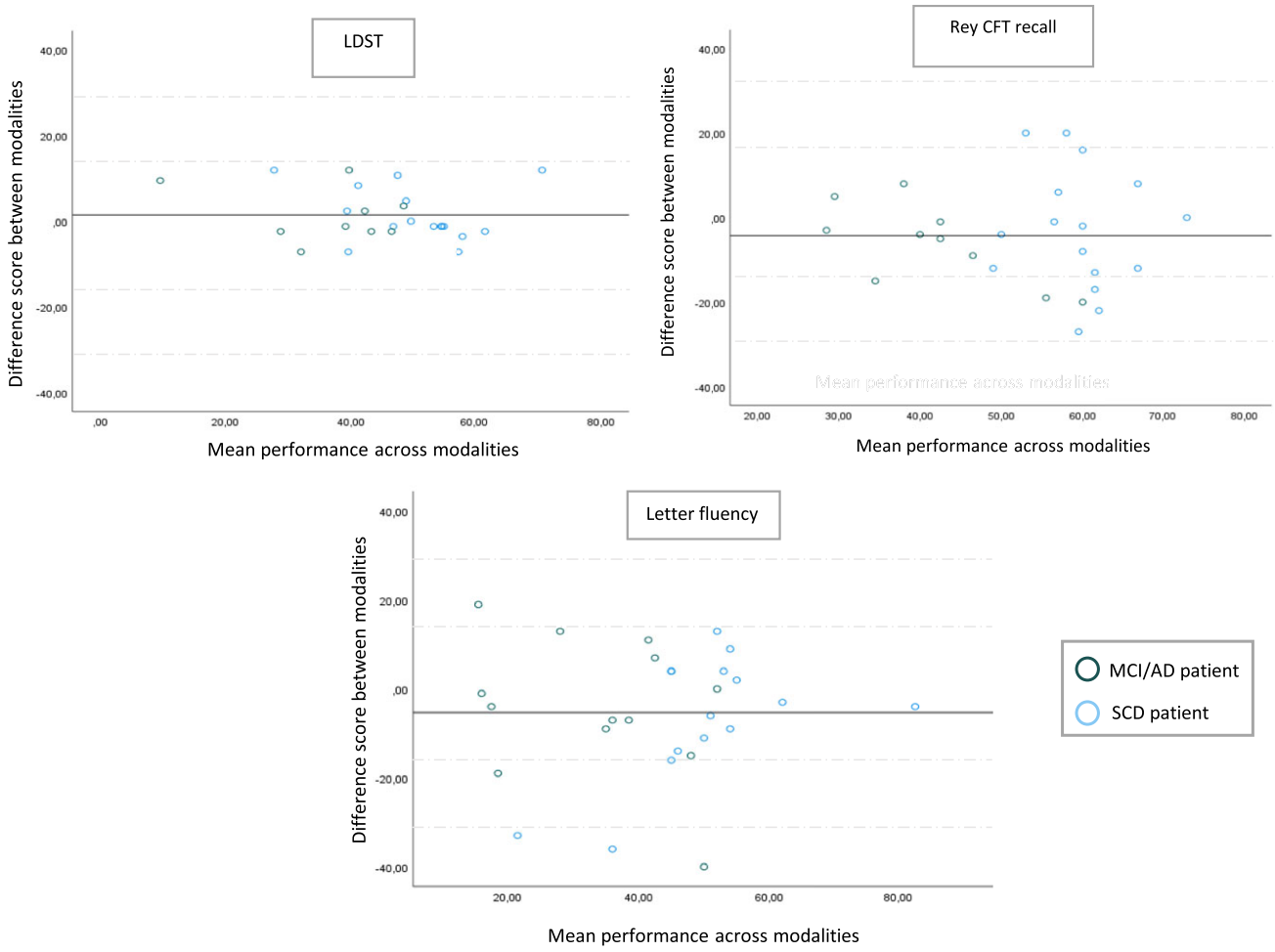

Figure 2. Bland–Altman plots showing test scores (T-scores), with the difference between measurements on the *Y*-axis and the mean of the two measurements on the *X*-axis. Dots represent individual patients (blue = Subjective Cognitive Decline, green = Mild Cognitive Impairment/dementia), the horizontal black line indicates mean difference on group-level between measurements, the dashed gray lines indicate ±1 and 2 SDs from a difference of $T = 0$. The three figures highlight tests that are highly stable across modalities (Letter Digit Substitution Test [LDST]), show apparent increasing variation between modalities for higher mean performances (Rey Complex Figure Test recall), and show apparent decreasing variation between modalities for higher mean performances (Letter Fluency).

with the highest number of patients showing a higher classification at baseline, while Rey CFT copy (34%), Digit Span Backward (27%) and Letter Fluency (21%) showed the highest number of patients with higher classification at VTC assessment.

VTC assessment – usability

Figure 3 shows total scores reported on the USE after the VTC assessment stratified by patient vs neuropsychologist reporting (a), and clinical diagnosis (b,c). Mann–Whitney U tests indicated no significant differences on total scores of the USE (p's > .05) as reported by patient or neuropsychologist for SCD vs MCI/dementia diagnosis. Supplementary figure 2 shows scores on the Ease of use and learning, Usefulness and Satisfaction domains. Neither the neuropsychologist- nor patient-reported mean scores on these domains differed for SCD vs MCI/dementia diagnosis.

Figure 4 shows the scores on each item of the SUS as judged by patients and neuropsychologists. As can be seen in the figure, the vast majority of patients agreed (i.e., responded with "somewhat agree" or "strongly agree") with positively formulated items, namely those stating that they would use the system frequently (84% agreed, item 1), that the VTC setup was easy to use (93% agreed, item 3) and that they felt confident using the VTC setup (81% agreed, item 9). Twenty-three percent of patients agreed that they would need help from someone with technical knowledge to use the VTC system (item 4). With regard to the patient-reported scores on the SUS, Fisher's exact tests indicated significant differences between SCD vs MCI/dementia patients in responses to item 4 (SCD patients had a higher proportion of "strongly disagree" response than MCI/dementia patients) and item 7 (SCD patients had a higher proportion of "somewhat agree" responses than MCI/dementia patients).

Similar to patients, most neuropsychologist responses were in agreement with items (i.e., responded with "somewhat agree" or "strongly agree") stating they would want to use the system frequently (87%, item 1) that the VTC setup was easy to use (83%, item 3), and that they felt confident using the VTC setup (77%, item 9). Fisher's exact tests revealed that neuropsychologists' scores on the SUS were significantly different for administrations concerning SCD vs MCI/dementia patients, specifically for item 8 (higher proportion of "strongly disagree" for SCD patients compared to MCI/dementia, and higher proportion of "somewhat disagree" for MCI/dementia patients compared to SCD).

In the focus group session, neuropsychologists reported technical issues as the biggest barrier, where delays in connection and trouble with audio were most prevalent. The home

https://doi.org/10.1017/S1355617724000432 Published online by Cambridge University Press

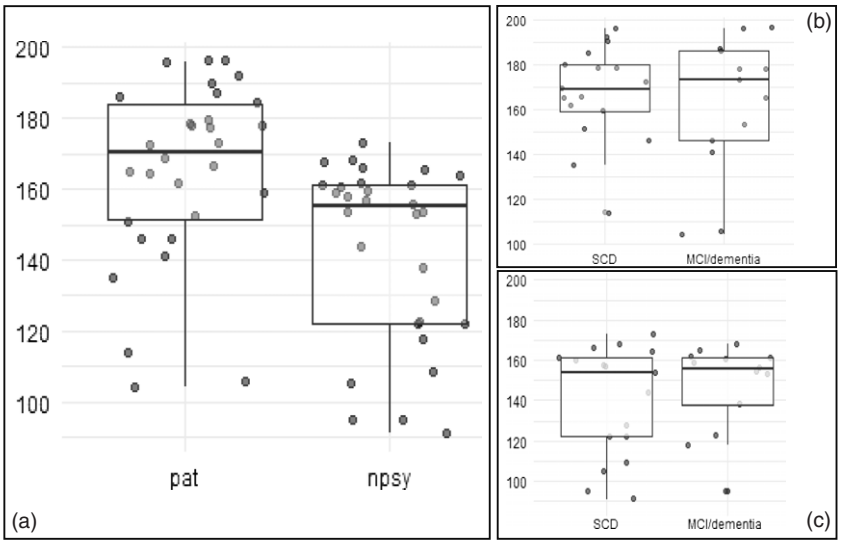

Figure 3. Total scores on the User Satisfaction and Ease of use questionnaire a) patient (pat) vs. neuropsychologist (npsy) reported scores, b) patient-reported scores stratified by clinical diagnosis (Subjective Cognitive Decline [SCD] vs Mild Cognitive Impairment [MCI]/dementia), and c) neuropsychologist reported scores stratified by clinical diagnosis (SCD vs MCI/dementia).

environment was perceived as more distracting compared to the consultation room, and it was harder to monitor and correct mistakes when needed (e.g., for the TMT). Clinical observations could be sufficiently made, and the neuropsychologists who touched on this topic mentioned that they did not experience differences in clinical observations during the VTC assessment compared to face-to-face. Some tests, including the VAT, were described as easier to administer digitally. In terms of patient suitability, neuropsychologists reported that patients without cognitive impairment (i.e., SCD diagnosis) were functioning more independent during the VTC assessment and needed less help with the technical setup of the VTC than patients with (extensive) impairment. An additional preparation session with the patient before the VTC assessment was helpful to prevent and timely resolve issues, e.g., with the MS Teams setup or with internet connection. Still, it was also mentioned that this added to workload.

Discussion

In this study, we aimed to add to the growing literature on reliability and usability of remote neuropsychological assessment by integrating a standard neuropsychological assessment protocol used for clinical diagnostic purposes in a Dutch memory clinic across face-to-face (baseline) and clinic-to-home VTC (follow-up) modalities in a repeated measures design. All of the test measures used in the workup provided sufficient to excellent reliability in the total patient sample. We observed stability of the clinical interpretation of the performances for most patients on all tests. Usability scores were positive, and most usefulness and usability indicators were rated similarly between patients with and without cognitive impairment (MCI/dementia vs SCD diagnosis).

Former studies have shown similar results regarding performances (Gnassounou et al., 2021; Vestal et al., 2006; Wadsworth et al., 2018), correlations (Galusha-Glasscock et al., 2016), and validity aspects (Martin-Khan et al., 2012) for face-to-face and VTC neuropsychological testing. Our results support and extend on this research by including a comprehensive test protocol that is administered in our memory clinic setting for clinical purposes as opposed to a selection of tests. The reliability findings (ICC or Pearson r/Spearman's rho) of the 24 test measures in the total sample were largely similar to those found in literature for repeated face-to-face assessment in adults (Bruijnen et al., 2020; Lezak et al., 2012; Schmand et al., 2008). For the RAVLT direct and delayed recall, standard MoCa and LDST, ICC values were somewhat higher. The fact that performances were overall more stable in in the MCI/dementia group than the SCD group and that SCD patients were mostly the ones showing better scores at the VTC follow-up is consistent with the notion that practice effects associated with repeated test administration tend to attenuate as cognition worsens (Jutten et al., 2020). We found no strong indication of a systematic difference in scores across modalities, except for the Rey Complex Figure test. Still, the correlation between the assessments' scores was similar to that of repeated face-to-face (Lezak et al., 2012).

Compared to previous findings in a design involving repeated VTC assessment of English speaking cognitively healthy adults (Fox-Fuller, Ngo, et al., 2022), it seems that reliability for verbal tests that tap into semantic fluency, attention and working memory (Fluency- Animals, Digit Span Forward and Backward tests) was somewhat higher in our study of older adults in the memory clinic. On the one hand, it may be beneficial for reliability if patients can familiarize themselves with neuropsychological tests in a face-to-face setting first, but we should also acknowledge that higher reliability could also be driven by other factors, such as the implementation of a VTC setup session with the neuropsychologist before actual assessment, and likely the high performance stability specifically present in our impaired patients.

It should be noted that reliability of some reaction time-based measures, such as Stroop test 1, was relatively low compared to other. This is not surprising, as they are granular scores that are particularly susceptible to intraindividual fluctuations from one assessment to the next, especially in elderly (Bielak et al., 2014). We hypothesize that these scores could be affected by subtle delays in connection during remote assessment, as reaction times at VTC follow-up were somewhat higher compared to face-to-face baseline. The neuropsychologists also reported timely correction of mistakes on the TMT remotely as more challenging, which could contribute to different completion times and accuracy. The amount of clinically relevant performances differences on these tests was, however, comparable to others.

A comprehensive evaluation of the applicability of a novel assessment modality requires evaluation of psychometric properties in conjunction with experiences of those who undergo and

https://doi.org/10.1017/S1355617724000432 Published online by Cambridge University Press

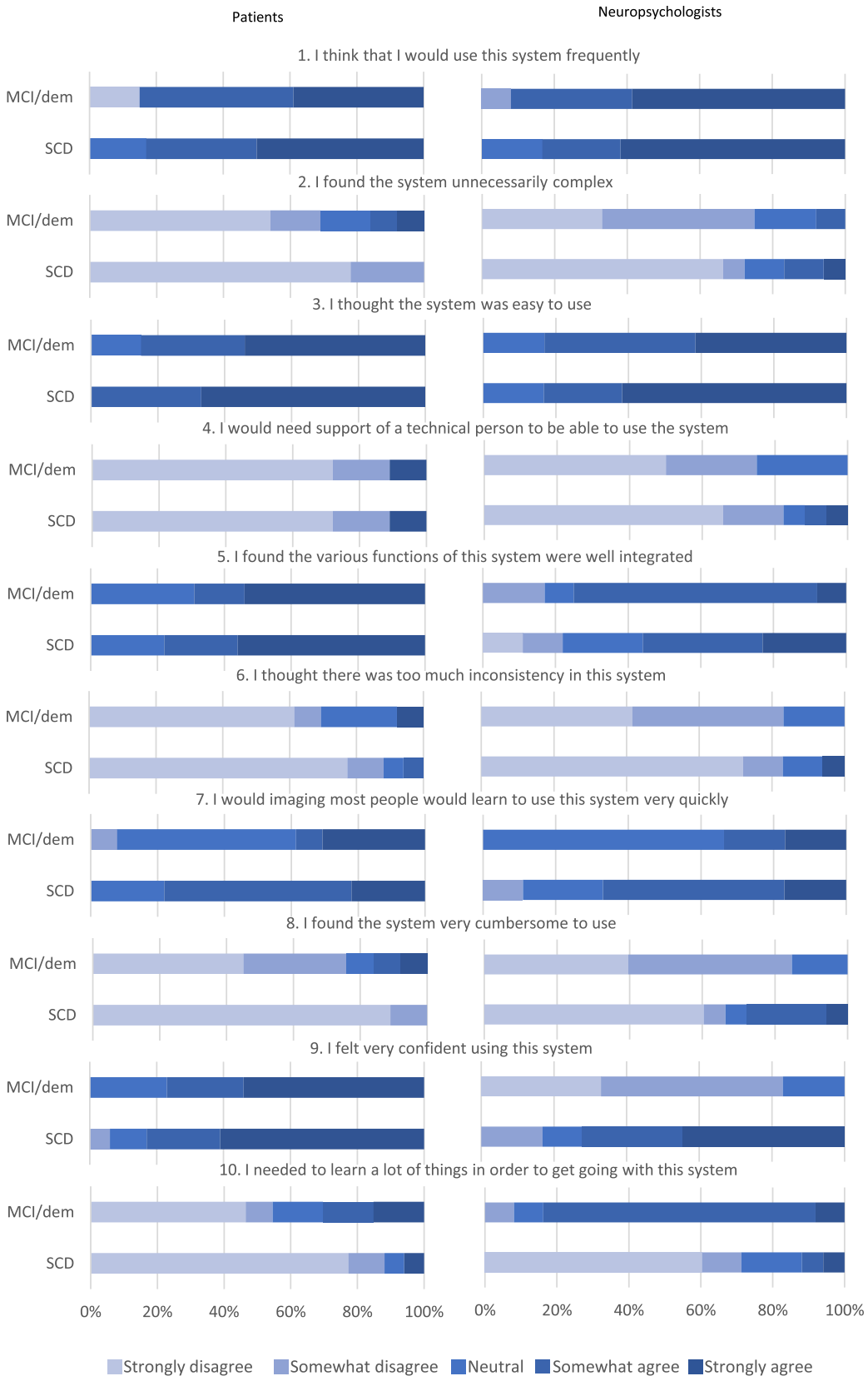

Figure 4. Responses on items of the System Usability Scale questionnaire for patients (left) and neuropsychologists (right).

administer it in daily practice (Parsons et al., 2022). Earlier research in cognitively unimpaired adults indicates patient and clinician satisfaction with remote testing in a Scottish neuropsychology service catering to a diverse (non-demented) patient population (Sumpter et al., 2023). While encouraging, this result cannot simply be generalized to the memory clinic setting where patients present with greatly varying degrees and types of functional and cognitive impairment that can influence ability

https://doi.org/10.1017/S1355617724000432 Published online by Cambridge University Press

to learn and use technology systems. Furthermore, one study found that face-to-face assessment was preferred more often in community-based individuals with MCI and Alzheimer's disease dementia than cognitively unimpaired individuals (Parikh et al., 2013).

We found positive system usability scores, and patients with and without cognitive impairment rated assessment with VTC equally in terms of usefulness, ease of use and satisfaction. Although the patient confidence in system usage corresponds with literature (Narasimh et al., 2017), we note that patients without cognitive impairment (SCD) were unsurprisingly more inclined to report confidence in their ability to learn the VTC system quickly than cognitively impaired patients (MCI/dementia diagnosis). More efficient learning of new skills in cognitively unaffected individuals should be taken into account (e.g., taking more time for system instructions and supervision if a patient has impairment). Interestingly, the neuropsychologists who performed the assessments rated similar user experiences for the diagnostic groups on the questionnaires, but they did report more problems and less independence in dementia patients in the focus group. Overall, their experience suggests that suitability for VTC assessment declines as cognitive and functional impairments progress over time.

Core experiences and challenges for US-based test administrators identified in previous survey research - having to adapt to a new mode of administration, occasional technological problems, environmental distractions in patients' homes, differences in technological fluency between individual patients – (Fox-Fuller, Rizer, et al., 2022) largely correspond with reporting from our neuropsychologists. In addition, our neuropsychologists reported that some tests were easier to administer through VTC, and that their ability for clinical observation during the assessment was largely unaffected. While additional preparations for VTC assessment in the form of a practice session with patients were seen as potentially helpful to prevent and timely resolve issues, it was also mentioned that this added to workload. The benefits and added efficiency for patients have to be weighed against such issues that can also result in additional strain on care planning and administration.

Decisions about when, how and for whom remote assessment in the memory clinic is a suitable option require in-depth evaluation (Bloch et al., 2021). The current study integrates some important issues that are key to such decisions, including reliability, clinical interpretability and experiences from key user groups. Although we should take caution in drawing definitive conclusions from this study alone due to the small sample size, we see a fairly consistent image across our and previous investigations that VTC administration can be done reliably (Fox-Fuller, Ngo, et al., 2022; Marra et al., 2020) and to users' satisfaction (Parsons et al., 2022) with some of the most commonly used neuropsychological tests, also in a memory clinic setting. It can be particularly useful for follow-up assessments of patients who may not have presented with formal cognitive impairment at a first visit, but should be monitored for timely identification of decline, as these patients may prefer remote testing more often and be more independent, requiring less instruction or help. This group could include individuals with amyloid positivity, subjective complaints that correspond with SCDplus, or a family history of degenerative disorders. Especially more advanced cases of dementia may not be suitable (Hunter et al., 2021).

Based on the results, it seems that tests that are timed, rely on speed, and/or are performed with pencil-and-paper can be administered through VTC with largely sufficient reliability and little deviation from repeated face-to-face administrations, suggesting that memory clinic administrators can test a broader range of functions beyond verbal (and) memory tests, such as attention, speed and cognitive flexibility. Naturally, administrating assessments remotely will require adjustments to regular care (Hewitt et al., 2022); the patient needs to receive assessment forms on time, possibly a practice session that will require time and personnel investment, and privacy concerns should be taken seriously. Furthermore, we should remain aware that in clinic-to-home assessments, some factors that can influence the assessment, but can be controlled with in-clinic assessments, are introduced .Such factors include external distractions and fluctuations in internet connections during the assessment.

It should be noted that reliability of neuropsychological testing can be investigated from different perspectives. Our design allows us to conclude that VTC assessment administered after initial in-person assessment has similar reliability as repeated in-person assessment, but it does not allow for a direct comparison of the modalities in absence of effects related to repeated testing, such as (varying) practice effects and test familiarity. In a heterogenous sample like ours consisting of patients with and without clinical diagnoses, these factors can vary. Still, this is inherent to all administration methods for neuropsychological assessment, not just remote testing. Studies with a counterbalanced design in which patients can receive either modality first or second would allowing for a direct comparison between different assessment modalities, and whether indeed reliability is better if the first assessment is done face-to-face. It is also important to address that we were able to include 31 patients with varied diagnoses in our inclusion period in the memory clinic. Similar studies that have been published with different samples are somewhat larger (Fox-Fuller et al., 2022; Wadsworth et al., 2016). Our results need to be replicated in a larger sample, which ideally would also be sufficient to subsequently more extensively investigate differences in reliability and user experiences as a function of socioeconomic and clinical characteristics to get a better picture of patient suitability. Importantly, our study was performed in as a tertiary diagnostic center in a Western country, and included a fairly high proportion of individuals with higher education with relatively good clinical status. Our sample therefore likely overrepresents access to and experience with technology of elderly individuals seeking help for (experienced) cognitive dysfunction, and may not reflect the entire memory clinic population. Finally, we note that due to a lack of normative data for tests administered through VTC assessment we used norms taken from in-person assessment, and that our cut off for clinically relevant change was based on clinical consensus.

In conclusion, reliability of remote clinic-to-home administration of tests of a standard neuropsychological workup were largely similar to repeated face-to-face assessment for most tests, with most patients showing no clinically relevant difference between modalities. Systematic and clinically relevant differences from face-to-face assessment appear limited. While we should remain aware of differences in technological access and literacy among elderly, more memory clinic patients, especially those at risk for cognitive decline in the future, may be candidate for VTC.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S1355617724000432.

Acknowledgements. Research of Alzheimer Center Amsterdam is part of the neurodegeneration research program of Amsterdam Neuroscience. Alzheimer Center Amsterdam is supported by Stichting Alzheimer Nederland and

Stichting Steun Alzheimercentrum Amsterdam. WF is supported by the Pasman stichting. SS and WF are recipients of TAP-dementia (www.tap-dementia.nl), receiving funding from ZonMw (#10510032120003) in the context of Onderzoeksprogramma Dementie, part of the Dutch National Dementia Strategy. No funding is provided for this specific study.

References

Altman, D. G., & Bland, J. M. (1983). Measurement in medicine: The analysis of method comparison studies. *Journal of the Royal Statistical Society. Series D (The Statistician)*, *32*(3), 307–317.

Appleman, E. R., O'Connor, M. K., Boucher, S. J., Rostami, R., Sullivan, S. K., Migliorini, R., & Kraft, M. (2021). Teleneuropsychology clinic development and patient satisfaction. *Clinical Neuropsychology*, *35*(4), 819–837.

Bielak, A. A., Cherbuin, N., Bunce, D., & Anstey, K. J. (2014). Intraindividual variability is a fundamental phenomenon of aging: Evidence from an 8-year longitudinal study across young, middle, and older adulthood. *Developmental Psychology*, *50*(1), 143–151.

Bilder, R. M., Postal, K. S., Barisa, M., Aase, D. M., Cullum, C. M., Gillaspy, S. R., Harder, L., Kanter, G., Lanca, M., Lechuga, D. M., Morgan, J. M., Most, R., Puente, A. E., Salinas, C. M., Woodhouse, J. (2020). InterOrganizational practice committee recommendations/guidance for teleneuropsychology (TeleNP) in response to the COVID-19 pandemic. *The Clinical Neuropsychologist*, *34*(7-8), 1314–1334.

Bloch, A., Maril, S., & Kavé, G. (2021). How, when, and for whom: Decisions regarding remote neuropsychological assessment during the 2020 COVID-19 pandemic. *Israel Journal of Health Policy Research*, *10*(1), 31.

Brearly, T. W., Shura, R. D., Martindale, S. L., Lazowski, R. A., Luxton, D. D., Shenal, B. V., & Rowland, J. A. (2017). Neuropsychological test administration by videoconference: A systematic review and meta-analysis. *Neuropsychology Review*, *27*(2), 174–186.

Brook, J. (1995). SUS: A quick and dirty usability scale. *Usability Evaluation In Industry*, *189*, 4–7.

Bruijnen, C., Dijkstra, B. A. G., Walvoort, S. J. W., Budy, M. J. J., Beurmanjer, H., De Jong, C. A. J., & Kessels, R. P. C. (2020). Psychometric properties of the Montreal Cognitive Assessment (MoCA) in healthy participants aged 18-70. *International Journal of Psychiatry in Clinical Practice*, *24*(3), 293–300.

Castanho, T., Amorim, L., Zihl, J., Palha, J., Sousa, N., & Santos, N. (2014). Telephone-based screening tools for mild cognitive impairment and dementia in aging studies: A review of validated instruments. *Frontiers in Aging Neuroscience*, *6*, 16. https://doi.org/10.3389/fnagi.2014.00016

Cullum, C., Hynan, L. S., Grosch, M., Parikh, M., & Weiner, M. F. (2014). Teleneuropsychology: Evidence for video teleconference-based neuropsychological assessment. *Journal of the International Neuropsychological Society*, *20*(10), 1028–1033.

Dorenkamp, M. A., Irrgang, M., & Vik, P. (2021). Assessment-related anxiety among older adults: Associations with neuropsychological test performance. *Sedentary Life and Nutrition*, *30*(2), 256–271.

Fox-Fuller, J. T., Ngo, J., Pluim, C. F., Kaplan, R. I., Kim, D.-H., Anzai, J. A. U., Yucebas, D., Briggs, S. M., Aduen, P. A., Cronin-Golomb, A., Quiroz, Y. T. (2022). Initial investigation of test-retest reliability of home-to-home teleneuropsychological assessment in healthy, English-speaking adults. *Clinical Neuropsychology*, *36*(8), 2153–2167.

Fox-Fuller, J. T., Rizer, S., Andersen, S. L., & Sunderaraman, P. (2022). Survey findings about the experiences, challenges, and practical advice/solutions regarding teleneuropsychological assessment in adults. *Archives of Clinical Neuropsychology*, *37*(2), 274–291.

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*, *31*(1), 8–11.

Geddes, M. R., O'Connell, M. E., Fisk, J. D., Gauthier, S., Camicioli, R., Ismail, Z., & the Alzheimer Society of Canada Task Force on Dementia Care Best Practices for COVID-19 (2020). Remote cognitive and behavioral assessment: Report of the Alzheimer Society of Canada Task Force on dementia care best practices for COVID-19. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *12*(1), e12111.

Gnassounou, R., Defontaines, B., Denolle, S., Brun, S., Germain, R. B., Schwartz, D., Schück, S., Michon, A., Belin, C., Maillet, D. (2021). Comparison of neuropsychological assessment by videoconference and face to face. *Journal of the International Neuropsychological Society*, *28*(5), 1–11.

Hendriks, M. B., & Kessels, R. P. C. (2020). Uniformiteit in de kwalitatieve beschrijving van scores op prestatietaken. *Tijdschrift voor Neuropsychologie*, *15*(3), 166–176.

Hewitt, K. C., Block, C., Bellone, J. A., Dawson, E. L., Garcia, P., Gerstenecker, A., Grabyan, J. M., Howard, C., Kamath, V., LeMonda, B. C., Margolis, S. A., McBride, W. F., Salinas, C. M., Tam, D. M., Walker, K. A., Del Bene, V. A. (2022). Diverse experiences and approaches to tele neuropsychology: Commentary and reflections over the past year of COVID-19. *The Clinical Neuropsychologist*, *36*(4), 790–805.

Hildebrand, R. H., Chow, H., Williams, C., Nelson, M., & Wass, P. (2004). Feasibility of neuropsychological testing of older adults via videoconference: Implications for assessing the capacity for independent living. *Journal of Telemedicine and Telecare*, *10*(3), 130–134.

Hunter, M. B., Jenkins, N., Dolan, C., Pullen, H., Ritchie, C., & Muniz-Terrera, G. (2021). Reliability of telephone and videoconference methods of cognitive assessment in older adults with and without dementia. *Journal of Alzheimer's Disease*, *81*(4), 1625–1647.

Jutten, R. J., Grandoit, E., Foldi, N. S., Sikkes, S. A. M., Jones, R. N., Choi, S-Eun, Lamar, M. L., Louden, D. K. N., Rich, J., Tommet, D., Crane, P. K., Rabin, L. A. (2020). Lower practice effects as a marker of cognitive performance and dementia risk: A literature review. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *12*(1), e12055.

Jutten, R. J., Harrison, J., Lee Meeuw Kjoe, P. R., Opmeer, E. M., Schoonenboom, N. S. M., de Jong, F. J., Ritchie, C. W., Scheltens, P., Sikkes, S. A. M. (2018). A novel cognitive-functional composite measure to detect changes in early Alzheimer's disease: Test-retest reliability and feasibility. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *10*(1), 153–160.

König, A., Zeghari, R., Guerchouche, R., Duc Tran, M., Bremond, Fçois, Linz, N., Lindsay, H., Langel, K., Ramakers, I., Lemoine, P., Bultingaire, V., Robert, P. (2021). Remote cognitive assessment of older adults in rural areas by telemedicine and automatic speech and video analysis: Protocol for a cross-over feasibility study. *BMJ Open*, *11*(9), e047083–e047083.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractice Medicine*, *15*, 155–163.

Lezak, M. D., H., D., Bigler, E. D., & Tranel, D. (2012). A compendium of tests and assessment techniques. In M. D. Lezak, D. H., E. D. Bigler, & D. Tranel (Eds.), *Neuropsychological assessment* (5th ed.). Oxford University Press.

Lunardini, F., Luperto, M., Romeo, M., Basilico, N., Daniele, K., Azzolino, D., Damanti, S., Abbate, C., Mari, D., Cesari, M., Borghese, N. A., Ferrante, S. (2020). Supervised digital neuropsychological tests for cognitive decline in older adults: Usability and clinical validity study. *Journal of Medical Internet Research*, *8*(9), e17963.

Lund, A. (2001). Measuring Usability with the USE Questionnaire. *Usability and User Experience Newsletter of the STC Usability SIG*, *8*, 3–6.

Marra, D. E., Hamlet, K. M., Bauer, R. M., & Bowers, D. (2020). Validity of teleneuropsychology for older adults in response to COVID-19: A systematic and critical review. *The Clinical Neuropsychologist*, *34*(7-8), 1411–1452.

Martin-Khan, M., Flicker, L., Wootton, R., Loh, P.-K., Edwards, H., Varghese, P., Byrne, G. J., Klein, K., Gray, L. C. (2012). The diagnostic accuracy of telegeriatrics for the diagnosis of dementia via Video Conferencing. *Journal of the American Medical Directors Association*, *13*(5), 487.e419–487.e424.

Narasimha, S., Madathil, K. C., Agnisarman, S., Rogers, H., Welch, B., Ashok, A., Nair, A., McElligott, J. (2017). Designing telemedicine systems for geriatric patients: A review of the usability studies. *Telemedicine and e-Health*, *23*(6), 459–472.

Owens, A. P., Ballard, C., Beigi, M., Kalafatis, C., Brooker, H., Lavelle, G., Brønnick, K. K., Sauer, J., Boddington, S., Velayudhan, L., Aarsland, D. (2020). Implementing remote Memory Clinics to enhance clinical care during and after COVID-19 [Hypothesis and Theory]. *Frontiers in Psychiatry*, *11*(990). https://doi.org/10.3389/fpsyt.2020.579934

Parikh, M., Grosch, M. C., Graham, L. L., Hynan, L. S., Weiner, M., Shore, J. H., & Cullum, C. M. (2013). Consumer acceptability of brief videoconference-

https://doi.org/10.1017/S1355617724000432 Published online by Cambridge University Press

based neuropsychological assessment in older individuals with and without cognitive impairment. *The Clinical Neuropsychologist*, *27*(5), 808–817.

Parks, A. C., Davis, J., Spresser, C. D., Stroescu, I., & Ecklund-Johnson, E. (2021). Validity of in-home teleneuropsychological testing in the wake of COVID-19. *Archives of Clinical Neuropsychology*. https://doi.org/10.1093/arclin/acab002

Parsons, M. W., Gardner, M. M., Sherman, J. C., Pasquariello, K., Grieco, J. A., Kay, C. D., Pollak, L. E., Morgan, A. K., Carlson-Emerton, B., Seligsohn, K., Davidsdottir, S., Pulsifer, M. B., Zarrella, G. V., Burstein, S. M., Mancuso, S. M. (2022). Feasibility and acceptance of direct-to-home tele-neuropsychology services during the COVID-19 pandemic. *Journal of the International Neuropsychological Society*, *28*(2), 210–215.

Pulsifer, M. B., Grieco, J. A., Burstein, S. M., Parsons, M. W., Gardner, M. M., & Sherman, J. C. (2021). The development and implementation of teleneuropsychology in an academic lifespan neuropsychology center: Lessons learned from the COVID-19 pandemic. *Journal of Clinical and Experimental Neuropsychology*, *43*(8), 774–785.

Rochette, A. D., Rahman-Filipiak, A., Spencer, R. J., Marshall, D., & Stelmokas, J. E. (2021). Teleneuropsychology practice survey during COVID-19 within the United States. *Applied Neuropsychology: Adult*, 1–11. https://doi.org/10.1080/23279095.2021.1872576

Schmand, B., Groenink, S. C., & van den Dungen, M. (2008). Letterfluency: Psychometrische eigenschappen en Nederlandse normen. *Tijdschrift voor Gerontologie en Geriatrie*, *39*(2), 64–74.

Singh, S., & Germine, L. (2021). Technology meets tradition: A hybrid model for implementing digital tools in neuropsychology. *International Review of Psychiatry*, *33*(4), 382–393.

Sumpter, R., Camsey, E., Meldrum, S., Alford, M., Campbell, I., Bois, C., O'Connell, S., Flood, J. (2023). Remote neuropsychological assessment: Acceptability and feasibility of direct-to-home teleneuropsychology methodology during the COVID-19 pandemic. *Clinical Neuropsychologist*, *37*(2), 432–447.

Tsiakiri, A., Koutzmpi, V., Megagianni, S., Toumaian, M., Geronikola, N., Despoti, A., Kanellopoulou, S., Arampatzi, X., Margioti, E., Davila, A., Zoi, P., Kalligerou, F., Liozidou, A., Tsapanou, A., Sakka, P. (2022). Remote neuropsychological evaluation of older adults. *Applied Neuropsychology: Adult*, *31*(5), 1–8.

Underwood, B. R., Thompsell, A., Sidhom, E., & Burns, A. (2021). Providing memory assessment services during COVID-19. *Aging & Mental Health*, *25*(8), 1373–1375.

van der Elst, W., van Boxtel, M. P., van Breukelen, G. J., & Jolles, J. (2006). The Letter Digit Substitution Test: Normative data for 1,858 healthy participants aged 24-81 from the Maastricht Aging Study (MAAS): Influence of age, education, and sex. *Journal of Clinical and Experimental Neuropsychology*, *28*(6), 998–1009.

van der Flier, W. M., & Scheltens, P. (2018). Amsterdam Dementia Cohort: Performing research to optimize care. *Journal of Alzheimer's Disease*, *62*(3), 1091–1111.

Vestal, L., Smith-Olinde, L., Hicks, G., Hutton, T., & Hart, J., Jr (2006). Efficacy of language assessment in Alzheimer's disease: Comparing in-person examination and telemedicine. *Clinical Interventions in Aging*, *1*(4), 467–471.

Wadsworth, H. E., Galusha-Glasscock, J. M., Womack, K. B., Quiceno, M., Weiner, M. F., Hynan, L. S., Shore, J., & Cullum, C. M. (2016). Remote neuropsychological assessment in rural American Indians with and without cognitive impairment. *Archives of Clinical Neuropsychology.*, *31*(5), 420–425.

Wadsworth, H. E., Dhima, K., Womack, K. B., Hart, J., Weiner, M. F., Hynan, L. S., & Cullum, C. M. (2018). Validity of teleneuropsychological assessment in older patients with cognitive disorders. *Archives of Clinical Neuropsychology*, *33*(8), 1040–1045.

World Medical Association (2013). World Medical Association declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA*, *310*(20), 2191–2194. https://doi.org/10.1001/jama.2013.281053

https://doi.org/10.1017/S1355617724000432 Published online by Cambridge University Press