An evaluation of the convergent validity of a face-to-face and virtual neuropsychological assessment counter balanced

Krynicki, Carl R.; Hacker, David; Jones, Christopher A.

DOI:
10.1111/jnp.12300

Received: 7 June 2022 | Accepted: 27 October 2022

DOI: 10.1111/jnp.12300

RESEARCH ARTICLE

the british psychological society
promoting excellence in psychology

An evaluation of the convergent validity of a face-to-face and virtual neuropsychological assessment counter balanced

Carl R. Krynicki[1,2] | **David Hacker**[3] | **Christopher A. Jones**[2]

[1]School of Psychology, The University of Birmingham, Birmingham, UK

[2]Birmingham and Solihull Mental Health NHS Foundation Trust, Birmingham, UK

[3]Clinical Neuropsychology Department, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK

Correspondence

Carl R. Krynicki, School of Psychology, University of Birmingham, 52 Pritchatts Road, Birmingham, UK.
Email: c.krynicki@bham.ac.uk

Abstract

The COVID-19 pandemic has highlighted the need for further research evaluating the validity of conducting a battery of neuropsychological assessments virtually compared with face-to-face administration. Previous research has suggested that some neuropsychological assessments yield valid results when administered virtually, however, much of the previous research focused on older adults. To determine the validity of virtually administered neuropsychological tests, 28 healthy participants were assessed using a within-subjects, counter-balanced design. Participants completed a neuropsychological assessment battery covering tests of general intellectual functioning, memory and attention, executive functioning, language and information processing speed, as well as effort. There was no significant difference between face-to-face administration of the neuropsychological battery compared with virtual administration for the majority of the tests used. However, there were significant differences in the Colour Naming Task, with participants making fewer errors on the colour naming task and inhibition/switching task when administered virtually compared with face-to-face administration. There was also a significant age cohort effect in the inhibition/switching task. There was also a trending significant difference in mode of administration for the Verbal Fluency Task. Virtually administered neuropsychological assessments largely provide a valid alternative to face-to-face assessments; however, consideration must be given to test selection as well as the population of participants that are being assessed. Other important considerations must focus on preserving the security and integrity of test materials, as well as administration in a medico-legal setting. Future research should focus on validating assessments with specific patient populations and developing a neuropsychological assessment battery using information technology.

KEYWORDS
cognition, neuropsychological assessment, Teleneuropsychology

INTRODUCTION

The use of videoconferencing software (telemedicine) has seen a rapid growth in recent years, in particular since the outbreak of the COVID-19 pandemic in March 2020 (World Health Organization, 2020). The pandemic resulted in an increase in the use of videoconferencing software for patient consultations and acutely highlighted the potential usefulness of the option to administer a neuropsychological assessment battery remotely where face-to-face contact is neither practical nor safe. Remote assessment also has the potential, added benefit of allowing individuals to remain in a setting that is comfortable to them for their neuropsychological assessment, which has been found to be favourable option for assessment, particularly for those in rural locations (Hilty et al., 2007; Norman, 2006), and those with limited mobility (Grosch et al., 2011).

The British Psychological Society (BPS) Division of Neuropsychology (DON) and the Inter Organisational Practice committee (Bilder et al., 2020) have supported the use of remote technology when undertaking clinical neuropsychological work and have made recommendations regarding the technical considerations that must be considered when conducting a neuropsychological assessment, such as appropriate internet speeds, access to technology and appropriate selection of neuropsychological assessments that may be easily performed online and yield comparable results if administered face to face (Bunnage et al., 2020). However, to date, the evidence for the equivalence of remote and face-to-face neurocognitive assessment has remained limited.

One of the unique aspects of neuropsychological assessment is the reliance on the use of well-standardised and reliable neurocognitive tests that are sensitive to cognitive impairment arising from a range of neurological conditions including moderate-to-severe TBI (Dikmen et al., 1995; Donders et al., 2015; Mazaux et al., 1997; Sigurdardottir et al., 2015), neurodegenerative conditions such as dementia (Bondi et al., 2009), stroke (Sinanović, 2010) and tumours (Wefel et al., 2018). Typically such batteries would incorporate not just broad screening assessments but more detailed and sensitive tests of executive functioning, memory, processing speed, working memory, attention and language. However, much of the literature on the equivalence of face-to-face and remote neurocognitive assessment has been limited to examining either broad screening tools [e.g. Mini Mental State Examination (MMSE)] or isolated tests of specific functions which do not adequately reflect the more extensive test batteries used by neuropsychologists.

Initial studies have been encouraging. Munro Cullum and Grosch (2013), for example assessed 83 adults with cognitive impairment and 119 healthy controls in a counterbalanced design to establish whether a video teleconferencing-based neuropsychological assessments yielded valid and reliable results compared with standard face-to-face administration. They focused on the MMSE (Folstein et al., 1975), Hopkins Verbal Learning Test—Revised (Benedict et al., 1998), Digit Span forward and backward (Wechsler, 2008), short form Boston Naming Test (Kaplan et al., 1983), Letter and Category Fluency (Delis et al., 2001) and Clock Drawing (Agrell & Dehlin, 1998). They concluded that administering these assessments using telecommunication software provided a valid alternative to face-to-face assessments. A study of 150 adults by Gnassounou et al. (2021) also found no difference between face-to-face and virtual administration of a brief neuropsychological assessment battery composed of the MMSE (Folstein et al., 1975), Free and Cued Selective Reminding Test (French version; Van der Linden et al., 2004), Mahieux Gestural Praxis Battery (Mahieux-Laurent et al., 2009), Frontal Assessment Battery (Dubois et al., 2000), Trail Making

Test (TMT; completion time and errors for part B; Godefroy et al., 2010), Rey–Osterreith Complex Figure (Meyers & Meyers, 1995) and Categorical and Phonological Verbal Fluency Tests (Godefroy et al., 2010). They did however identify small significant differences in the Digit Span Forwards and Backwards (Wechsler, 2008) and the number of errors on the TMT part A (Godefroy et al., 2010).

Moreover, a meta-analytic review of 12 papers ($n = 497$) by Brearly et al. (2017) provided support for the use of video conferencing software to administer a neuropsychological assessment to a heterogenous clinical sample of patients (such as dementia, mild cognitive impairment, mixed sample of neurological disease and healthy individuals), especially assessments that rely on verbal responses from participants. However, only two of the studies included in this meta-analysis drew from a sample of individuals below the age of 65, therefore questioning how transferrable these findings may be to the working age, adult population. Furthermore, many of the included studies focused on basic screening tools such as the MMSE (Loh et al., 2007; Montani et al., 1997) which may be relatively insensitive to detecting neurocognitive deficits. There is, therefore, a pressing need for studies assessing the how comparable scores obtained from remote testing are compared with face-to-face assessment, using a more comprehensive neurocognitive test battery.

Rationale and aim

Typically, the performance of individuals with suspected cognitive impairment is compared with that of a normative sample representative of that individual's characteristics, typically age, gender and education. In the present study, the authors assess the equivalence of scores obtained by a normative sample administered a relatively comprehensive neurocognitive test battery, using counterbalanced face-to-face and remote assessment.

Predictions and hypotheses

In light of the previous research findings, a null hypothesis which predicted that there will be no difference between administering a battery of neuropsychological assessments virtually compared with face-to-face administration was adopted.

METHODOLOGY

Design

A counterbalanced, within-subject's experimental design was utilised. Participants were assessed twice, once face-to-face and once using telecommunication software (such as Zoom or Microsoft Teams). Participants were required to complete the assessment on a computer or laptop with a minimum screen size of thirteen inches. The study was counterbalanced as half of the participants firstly completed the neuropsychological assessment battery face to face, and the other half being administered the neuropsychological assessment battery using telecommunication software first. This was to reduce any confound caused by repeated exposure to the test stimuli ('practice effects').

Recruitment and procedure

Eligible participants were recruited via online advertisements, and participants were randomly allocated using computer software to receive either the face-to-face or the virtual neuropsychological assessment battery first. Participants were not incentivised for taking part. The assessment battery took around 2 h to complete, after which a suitable time to complete the second part of the assessment (either virtual or face-to-face depending on the mode of their first assessment) was agreed upon with a minimum period between assessments being 1 day.

Inclusion/exclusion criteria

Based on the normative data available for tests in the assessment battery, adults aged between 18 and 89 eligible to take part in the study. Additionally, those without a diagnosis of a neurological condition or learning disability, who were English speakers (to a sufficient standard that it would not invalidate the standard administration of the test) and able to give informed consent were eligible to take part in the study. Participants also needed access to the internet, the ability to use telecommunication software (either Zoom, Skype or Microsoft Teams) and access to an appropriate screen size.

Materials

The neuropsychological assessment battery was chosen to provide an overview of an individual's cognitive functioning.

Motivational and effort testing

Participants performance validity was assessed using a standalone performance validity test (PVT): The Test of Memory Malingering (TOMM), trial one (Tombaugh, 1996). The TOMM has been shown to be both a valid and reliable measure of effort (Martin et al., 2020). In addition to the stand-alone PVT, there were embedded PVT's within some of the neuropsychological assessments. The Reliable Digit Span score of the Wechsler Adult Intelligence Scale 4th Edition (WAIS-IV; Wechsler, 2008) Digit Span was used and has been found to have 93% specificity when set at a cut-off of less than seven (Schroeder et al., 2012). Moreover, the recognition condition of the Wechsler Memory Scale 4th Edition (WMS-IV; Wechsler, 2009) Logical Memory test and Visual Reproduction test were used to assess effort and have been found to have modest sensitivity and high specificity (Bouman et al., 2016; Soble et al., 2019).

In order to assess a participants effort level, one stand-alone and three embedded measures of test validity were used. Participants were excluded if they scored below the established cut-off score of 42 on the TOMM (Trial 1; Martin et al., 2020) and two of the embedded effort measures in one assessment. Table 1 shows the cut-off score for the embedded and standalone effort measures.

Based on this criterion, none of the participants included in this study scored below 42 on the TOMM and failed two of the embedded effort measures; therefore, none of the participants' data was removed from the analysis.

General intellectual functioning

Premorbid functioning

The Test of Premorbid Functioning—United Kingdom (TOPF^{UK}; Wechsler, 2008) was utilised to assess premorbid functioning. The TOPF^{UK} consists of seventy words that increase in unfamiliarity and

TABLE 1 Cut-off scores for the embedded effort measures

Test	Cut-off score	References
Logical memory (recognition)	15	Holdnack et al. (2013)
Visual reproduction (recognition)	3	Holdnack et al. (2013)
Reliable Digit Span	7	Holdnack et al. (2013)

irregularity. This test helps to determine an approximate premorbid level of functioning. The TOPF^{UK} manual not only reports excellent internal consistency (r = .92–.99) and test–retest reliability (r = .89–.95), but the TOPF^{UK} also correlates with the other WAIS-IV measures of general intellectual functioning (r = .70) and verbal intelligence (r = .75; Holdnack & Drozdick, 2009).

Current intellectual function

The WAIS-IV (Wechsler, 2008) Information and Matrix Reasoning sub-tests were used to assess current intellectual functioning. The matrix reasoning subtest and information subtests are relatively resistant to neurological insult such as severe TBI (Carlozzi et al., 2015) and can also give a guide to premorbid intellect.

Memory and attention functioning

Participants were administered two subtests from the WMS-IV (Wechsler, 2009): a test of verbal episodic memory (Logical Memory) and a test of recall for non-verbal visual stimuli (Visual Reproduction) both of which test immediate recall and delayed recall after 20–30 min as well as delayed recognition (included in this instance as embedded validity measures). The WMS-IV has been found to provide a valid and reliable measure of memory (Lo et al., 2012).

The Digit Span sub-test of the WAIS-IV (Wechsler, 2008) was used to assess attention, concentration, verbal and working memory and provides a reliable and valid measure of intellectual functioning (Girard et al., 2014).

The Rey–Osterreith Complex Figure task (Osterrieth, 1944; Rey, 1941) is a measure of constructional ability, perceptual–organisational ability and visual-perceptual memory. Participants were asked to copy the complex figure and then reproduce the figure from memory 3 min later (immediate recall) and 30 min later (delayed recall). Between each condition of the Rey–Osterreith Complex Figure and the visual reproduction task of the WMS, participants were asked to hide the previous drawing and were instructed to take a photograph of each drawing and send them to a member of the research team.

Executive functioning

The Oral Trail Making Test (Ricker & Axelrod, 1994) assesses attention, tracking and maintenance of cognitive set-shifting. Participants were first asked to count from one to twenty-five as quickly as possible, but without making any errors. For the second condition, participants were required to alternate between numbers and letters as quickly as possible without making any errors. The normative data used were extracted from Mrazik et al. (2010).

Two sub-tests of the Delis–Kaplan Executive Function System test battery (Delis et al., 2001) were administered. The Verbal Fluency test was used as it evaluates the spontaneous cognitive initiation, set-shifting and cognitive flexibility while under restricted search conditions. Participants completed the letter fluency, category fluency and category switching trials. The Colour–Word Interference Test was used to measure the participants ability to maintain a goal and suppress a habitual response. Participants completed the colour naming, word reading, colour word interference (inhibition) and colour word switching (attentional inhibition and switching) trials.

The Hayling Sentence Completion test and the Brixton Spatial Anticipation Tests (Burgess & Shallice, 1997) were used to evaluate initiation speed and response suppression as well as rule learning and cognitive flexibility. The Hayling Sentence Completion Test and Brixton Spatial Anticipation Test have been found to reliably and sensitivity measure frontal lobe functions (Bagshaw et al., 2014; Robinson et al., 2015).

Information processing speed

The Oral Symbol Modalities Test (Smith, 1973) was used to assess information processing speed, divided attention, visual scanning and tracking. Participants were presented with nine symbols with a corresponding number, presented in a legend above the test items. Participants were then required to pair the number that corresponded to a symbol as quickly as possible within 90 s. The test is similar to the coding task from the WAIS-IV but requires only a verbal response. Normative data from Strober et al. (2020) were utilised.

Language functioning

The full version of the Boston Naming Test (Kaplan et al., 1983) was used to assess language function and consists of a confrontation naming test utilising 60 images, named either spontaneously or with semantic or phonemic cues.

In addition to the neuropsychological assessment battery, participants were also asked to provide demographic information consisting of their date of birth, ethnicity, gender and number of years in education.

Ethical approval and issues

Ethical approval was sought and received from the University of Birmingham Research Ethics Committee (ERN_21-1412). A comprehensive risk assessment with the School of Psychology was also carried out, assessing the additional risk of face-to-face assessments during the COVID-19 pandemic. Approval from the School of Psychology's Risk Assessment Committee was also sought and received (RA SOPHS_21_100_CJ).

Statistical analysis

A database using Statistical Package for the Social Sciences (SPSS v.22) was created to manage the statistical analyses. All analyses carried out were of one-tailed significance unless otherwise stated and the Alpha level was set at .05. Descriptive statistics were explored, and an analysis of covariance (ANCOVA) model was created for each variable (face-to-face and its corresponding virtual variable), using the raw scores from each assessment (except for the Hayling Sentence Completion Test and Brixton Spatial Anticipation Test which were analysed using Scaled Scores), controlling for age, the date of the first assessment and days between the assessments. These were entered as covariates to control for these variables in the analysis to determine whether there were any differences between administering the neuropsychological assessment battery online to virtual administration. Data were also subjected to bootstrapping, which resamples a dataset to create simulated samples, therefore overcoming the limitation of smaller sample sizes.

Finally, the two one-sided test approach to equivalence testing was carried out using the TOSTER function of the Jamovi (v. 2.3.18) statistical package to assess whether the effect of the differences between face-to-face and virtual administration of the test material was meaningful.

RESULTS

Demographics and descriptive statistics

Twenty-eight healthy participants were recruited and assessed face-to-face and virtually with a battery of commonly used psychometric tests. The mode of administration (i.e. face-to-face or virtual administration) was counterbalanced across participants. The order of the psychometric tests within a mode of administration was the same for all participants.

The demographic and descriptive statistics of participants can be found in Table 2.

As there was a significant difference in the length of time between first and second assessment ($t = -3.035$, $p = .005$), the length of time between first and second assessment will be used as a covariate in the subsequent analysis of discrepancies between face-to-face and virtual test administration.

TABLE 2 Demographics and descriptive statistics

	Male			Female		
	Count	Mean	Standard deviation	Count	Mean	Standard deviation
Ethnicity[a]						
White	11			14		
Asian	0			2		
Mixed	0			1		
Age[b]	11	41.27_a	15.15	17	39.12_a	12.14
Years in education[b]	11	14.09_a	2.88	17	15.06_a	2.30

[a]Chi-square test of sex by ethnicity = $X^2 = 2.174$, $p = .337$, exact $p = .505$.

[b]Values in the same row not sharing the same subscript are significantly different at $p < .05$ in the two-sided test of equality for column means. Cells with no subscript are not included in the test. Tests assume equal variances.

TABLE 3 Significant difference in the length of time between first and second assessment

	First assessment					
	Face to face			Virtual		
	Mean	Standard deviation	Count	Mean	Standard deviation	Count
Days between assessments	24_a	18	14	50_b	26	14

Note: Values in the same row and sub-table not sharing the same subscript are significantly different at $p < .05$ in the two-sided test of equality for column means. Cells with no subscript are not included in the test. Tests assume equal variances.

Time differences between face-to-face and virtual administration

The difference in the length of time between first and second assessment is shown in Table 3.

As there was a significant difference in the length of time between first and second assessment ($t = -3.035$, $p = .005$) in those participants that received the face-to-face assessment first compared with those who firstly received the virtual assessment, the length of time between first and second assessment will be used as a covariate in the subsequent analysis of discrepancies between face-to-face and virtual test administration. This covariate is included because it is plausible that practice effects may vary as a function of the length of time since the initial test administration.

Discrepancies between face-to-face and virtual test administration

The discrepancy between face-to-face and virtual test administration was assessed using a four-way ANCOVA. The within-subject factor was the mode of administration (face-to-face vs. virtual administration) and the between-subjects' factor was the counterbalancing of the initial mode of administration.

Two covariates were also included. As previously noted, the average number of days between first and second test administration was significantly different depending on whether the first administration of the test was face-to-face or virtual. As the length of time between administrations could influence practice effects then the length in time (in days) between test administrations was included as a covariate. The second covariate was the age (in years) of the participant. This was included as some of the psychometric tasks show an ageing profile it is possible that raw scores may be influenced by the age of the participant.

The age and administration delay covariates appearing in the ANCOVA model were evaluated at their average values for the participants undertaking testing (age = 39.96 years and administration delay = 36.71 days).

Motivational and effort testing

In order to assess which mode of administration resulted in differences in performance on the TOMM (trial one), a four-way ANCOVA was constructed as indicated above.

TABLE 4 Dedicated and embedded measures of test validity

	Face-to-face administration		Virtual administration		Age cohort effect	First administration effect	Days between administration effect	Administration effect
	Mean	*SD*	Mean	*SD*	*F* (*p*)	*F* (*p*)	*F* (*p*)	*F* (*p*)
TOMM Trial 1	48.179	.457	48.607	.313	.073 (.789)	.555 (.464)	.007 (.934)	< .001 (.993)
Logical memory (recognition)	25	3.63	26.18	2.97	.023 (.880)	1.106 (.304)	1.190 (.286)	.435 (.516)
Visual reproduction (recognition)	6.32	.941	6.64	.488	1.651 (.211)	.174 (.680)	1.842 (.187)	.003 (.960)
Reliable digit span	9.14	2.050	9.43	2.348	.743 (.397)	.015 (.903)	.126 (.726)	.110 (.744)

There was no significant difference between the face-to-face and virtual administration of the TOMM ($F = <.001$, $p = .993$). Similarly, order of administration, the delay between administration and the age of participants did not affect performance on this task. A four-way ANCOVA was also constructed to assess performance on embedded measures of test validity differed by mode of administration. The tests were the recognition task of the Logical Memory and Visual Reproduction tasks within the WMS—IV (Wechsler, 2009) and the Reliable Digit Span (Holdnack & Drozdick, 2009) calculated from the Digit Span test within the WAIS—IV (Wechsler, 2008).

There was no significant difference between the face-to-face and virtual administration on any of the embedded measure of test validity (Logical Memory, $F = .435$, $p = .516$; Visual Reproduction, $F = .003$, $p = .960$; Reliable Digit Span, $F = .110$, $p = .744$; Holdnack & Drozdick, 2009; Wechsler, 2009). Similarly, order of administration, the delay between administration and the age of participants did not affect performance on this task. The results from the motivational and effort testing can be found in Table 4.

Performance on the face-to-face and virtually administered neuropsychological assessment

The results of the Test of Premorbid Functioning—United Kingdom (TOPF^{UK}; Wechsler, 2008), current intellectual functioning (measured using the WAIS-IV Information and Matrix Reasoning; Wechsler, 2008), memory and attention functioning [measured by WMS-IV (Wechsler, 2009), Digit Span from the WAIS-IV (Wechsler, 2008) and the Rey–Osterreith Complex Figure (Osterrieth, 1944; Rey, 1941)], executive functioning [assessed using The Verbal Fluency and Colour Word Interference Tests of the DKEFS test battery (Delis et al., 2001), the Oral Trail Making Test (Ricker & Axelrod, 1994) and the Hayling and Brixton Tests (Burgess & Shallice, 1997)], processing speed [measured using the Oral Symbol Modalities Test (Smith, 1973)] and language function [assessed using the Boston Naming Test (Kaplan et al., 1983)] can be found in Table 5. Data were analysed using a four-way ANCOVA, which was constructed to assess differences in mode of administration, and data were subject to bootstrapping.

There was a significant difference between face-to-face and virtual administration on the colour naming task (Delis et al., 2001). Participants performed significantly better in correcting an error made in the colour naming task when they were administered the virtual test than they did face to face, accounting for order of administration (Cohens $d = .19$). There was also an age cohort effect on the corrected errors on the colour naming task (Cohens $d = .44$), uncorrected errors on the colour naming task (Cohens $d = .19$) and the uncorrected errors on the inhibition/switching task (Cohens $d = .03$; Delis et al., 2001). Moreover, there was also a trending significant difference between the inhibition/switching uncorrected errors score, with participants making fewer errors on the virtual administration of the task compared to face-to-face administration. There was also a significant effect of delay between the follow-up assessment on the word reading uncorrected score and a trending significant effect of delay between the follow-up

TABLE 5 Performance on the face-to-face and virtually administered neuropsychological assessment

	Face-to-face administration		Virtual administration		Age cohort effect	First administration effect	Days between administration effect	Administration effect	Equivalence test	
	Mean	*SD*	Mean	*SD*	*F* (*p*)	*F* (*p*)	*F* (*p*)	*F* (*p*)	*t*	*p*
Test of premorbid functioning										
TOPF^{UK}	42.61	13.11	43.51	11.70	.937 (.343)	4.697 (.040)	.323 (.575)	.562 (.461)	.857	.399
Current intellectual functioning										
Information	17	4.56	17	4.51	1.587 (.220)	.005 (.947)	1.820 (.190)	.157 (.695)	.00	1.000
Matrix reasoning	18.11	4.67	17.36	4.066	1.477 (.236)	.023 (.880)	.869 (.361)	2.163 (.154)	.958	.347
Memory and attention functioning										
WMS IV logical memory										
Immediate	29.79	8.006	32.50	8.677	.176 (.678)	2.115 (.159)	2.001 (.170)	1.846 (.187)	2.08	.047
Delayed	27.36	8.385	28.36	8.341	.305 (.586)	23.009 (<.001)	1.955 (.175)	1.540 (.227)	.823	.418
WMS-IV visual recognition										
Immediate	38.68	4.730	38.61	4.856	.240 (.628)	7.366 (.012)	.167 (.686)	.055 (.817)	.131	.897
Delayed	35.04	7.234	35.43	5.534	.896 (.353)	9.720 (.005)	.120 (.732)	.573 (.457)	.326	.746
WAIS-IV digit span										
Forwards	10.18	2.568	10.61	2.601	.468 (.500)	4.203 (.051)	2.845 (.105)	.757 (.393)	1.400	.173
Backwards	7.00	2.108	6.50	2.301	1.758 (.179)	.349 (.560)	1.244 (.276)	.662 (.424)	1.240	.226
Sequencing	8.50	1.202	8.57	1.752	1.003 (.327)	1.569 (.222)	3.419 (.077)	.002 (.965)	.223	.826
Total	25.68	5.004	25.68	5.538	2.983 (.097)	.000 (.995)	.681 (.417)	.906 (.351)	.000	1.000
Rey–Osterreith complex figure										
Copy	34.32	2.310	34.18	2.881	2.114 (.159)	.115 (.737)	.277 (.604)	.832 (.371)	.202	.841
Immediate	23.089	7.789	23.125	7.069	.190 (.666)	3.459 (.075)	.022 (.883)	.067 (.798)	.023	.981
Delayed	21.946	7.754	22.679	7.596	.190 (.667)	9.713 (.005)	.161 (.691)	.000 (.992)	.584	.564
Executive functioning										
Hayling sentence completion (time)	46.21	32.004	59.18	32.740	.004 (.949)	2.556 (.123)	5.436 (.028)	.167 (.686)	1.224	.231
Brixton spatial anticipation test (errors)	16.86	9.236	15.86	6.311	3.136 (.089)	5.626 (.026)	.050 (.825)	1.543 (.226)	.711	.483
DKEFS colour–word interference test										

TABLE 5 (Continued)

	Face-to-face administration		Virtual administration		Age cohort effect	First administration effect	Days between administration effect	Administration effect	Equivalence test	
	Mean	*SD*	Mean	*SD*	*F* (*p*)	*F* (*p*)	*F* (*p*)	*F* (*p*)	*t*	*p*
Uncorrected colour naming	.00	.00	.14	.448	5.619 (.026)	2.876 (.103)	.533 (.472)	2.723 (.112)	1.686	.103
Corrected colour naming	.32	.25	.25	.441	5.559 (.027)	.283 (.599)	.733 (.401)	4.912 (.036)	.570	.573
Colour naming time	28.07	4.906	28.32	6.377	.840 (.368)	11.406 (.002)	4.225 (.051)	.046 (.833)	.298	.768
Word reading uncorrected	.00	.000	.04	.189	.296 (.592)	.188 (.669)	7.029 (.014)	1.500 (.233)	1.00	.326
Word reading corrected	.07	.262	.11	.315	.038 (.847)	.057 (.814)	.230 (.636)	.048 (.828)	.570	.573
Word reading time	21.96	5.847	21.54	4.887	.060 (.809)	1.679 (.207)	.179 (.676)	.011 (.916)	.450	.657
Inhibition uncorrected	.36	1.062	.25	.645	.175 (.680)	.757 (.393)	2.719 (.112)	.207 (.653)	.500	.621
Inhibition corrected	.68	1.056	.75	1.143	2.131 (.157)	1.891 (.182)	2.010 (.169)	2.475 (.129)	.338	.738
Inhibition time	52.18	15.435	52.57	14.992	.323 (.575)	.589 (.450)	8.020 (.009)	2.198 (.151)	.281	.780
Inhibition/switching uncorrected	.54	1.036	.50	1.00	6.258 (.020)	2.436 (.132)	.182 (.674)	3.930 (.059)	.171	.865
Inhibition/switching corrected	.68	1.020	1.14	1.484	.006 (.938)	.008 (.930)	.400 (.533)	.011 (.919)	1.455	.157
Inhibition/switching time	64.50	31.233	58.79	13.362	.835 (.370)	2.311 (.142)	3.085 (.092)	.078 (.782)	1.14	.264
DKEFS verbal fluency test										
Letters	43.82	10.951	44.11	12.764	3.980 (.058)	8.659 (.007)	.982 (.332)	3.287 (.082)	.237	.814
Category	50.75	7.457	51.75	8.081	.015 (.905)	2.607 (.119)	.077 (.784)	.002 (.965)	.732	.470
Switching	16.32	2.932	16.96	2.795	.562 (.461)	11.650 (.002)	1.584 (.220)	.567 (.459)	1.224	.231
Oral trail making test time	36.21	18.17	40.00	21.15	.067 (.797)	2.211 (.150)	.200 (.659)	.0095 (.760)	1.54	.136
Oral trail making test set loss errors	.21	.499	.43	.742	.457 (.506)	1.387 (.251)	1.238 (.277)	1.444 (.241)	1.24	.227
Oral trail making test sequential errors	.82	1.467	.71	.976	.131 (.720)	1.645 (.212)	.246 (.625)	.290 (.595)	.350	.729
Processing speed										
Oral symbols modalities test	56.07	15.639	57.79	13.248	1.323 (.261)	7.670 (.011)	2.350 (.138)	1.445 (.241)	1.110	.227
Language function										
Boston naming test	55.61	3.247	55.82	3.486	1.088 (.307)	7.440 (.012)	1.617 (.216)	.015 (.903)	.331	.743

assessment on the time to complete the colour naming task. A trending significant difference was found in the letter fluency condition of the DKEFS verbal fluency test (Delis et al., 2001; Cohens $d = .02$).

Order of presentation of TOPF^{UK} (Wechsler, 2008), the Delayed Logical Memory task and the Delayed Visual Recognition task (Wechsler, 2009), the letters and switching condition of the DKEFS Verbal Fluency task (Delis et al., 2001), colour naming time (Delis et al., 2001), Brixton Spatial Anticipation task (Burgess & Shallice, 1997), Oral Symbol Modalities Test (Smith, 1973) and the Boston Naming Test (Kaplan et al., 1983) did have a significant effect, whereby face-to-face administration scores appear lower in those who received the face-to-face administration first, whereas virtual administration scores appeared higher in those who received the virtual administration first compared with those who received virtual administration second.

Equivalence testing revealed that the majority of virtually and face-to-face administered neuropsychological tests yielded equivalent results, with the exception of the immediate condition of the WMS-IV Logical Memory test (Wechsler, 2009) which was not equivalent.

DISCUSSION

Summary of findings

The aim of this study was to explore whether there are differences between face-to-face and virtual administration of a battery of neuropsychological assessments. This study found that there was no significant difference between many of the neuropsychological assessments used in this study, which supports previous research exploring differences in performance in virtual and face-to-face neuropsychological assessments. There were no significant differences in mode of administration for the tests assessing motivation and effort and there were no significant differences between virtual and face-to-face administration of the WMS-IV (Wechsler, 2009) tests (Logical Memory and Visual Reproduction), the WAIS-IV (Wechsler, 2008; Test of Premorbid Functioning, Information, Matrix Reasoning and Digit Span) or in any of the three conditions of the Rey–Osterreith Complex Figure (Osterrieth, 1944; Rey, 1941). Regarding the tests of executive functioning, there was no significant difference in mode of administration for the Hayling and Brixton tests (Burgess & Shallice, 1997) or the Oral Trail Making Test (Ricker & Axelrod, 1994). Participants also did not differ in their performance on the Oral Symbol Modalities Test and Boston Naming Test (Kaplan et al., 1983) when administered face-to-face compared with virtual administration. The use of equivalence testing supported the finding that face-to-face and virtual administration of most of the neuropsychological tests were statistically equivalent, with the exception of the immediate condition of the Logical Memory test of the WMS-IV (Wechsler, 2009) which found that the virtual and face-to-face administration were not equivalent.

The results from this study identified significant differences in the DKEFS (Delis et al., 2001) corrected errors score on the colour naming task, with those being administered the task virtually performing significantly better than when completing the task face-to-face. However, the effect size of this difference is small, therefore suggesting that despite being statistically significant, the difference may not be clinically significant. There was also a trend for greater errors in the inhibition/switching uncorrected error score for the DKEFS (Delis et al., 2001) Colour–Word Interference Task, that did not reach statistical significance. Moreover, there was a trend for participants to produce more words in the D-KEFS Letter Fluency task (Delis et al., 2001), when administered remotely compared with face-to-face administration, but this did not reach significance.

The absence of a significant relationship between many of the neuropsychological assessments used in this study, which supports previous research exploring differences in performance in virtual and face-to-face neuropsychological assessments. Munro Cullum and Grosch (2013) found that there was no statistically significant difference in performance of the DKEFS Category Fluency (Delis et al., 2001), Boston Naming Test (Kaplan et al., 1983) and Digit Span (Wechsler, 2008) when administered virtually compared with face-to-face administration, which was supported by the present research. Moreover, there was a lack of significant associations between mode of assessment and performance on the Trail Making Test (Delis et al., 2001), Rey–Osterreith Complex Figure (Osterrieth, 1944; Rey, 1941) and the DKEFS Category Fluency test (Delis et al., 2001), which supports the findings by Gnassounou et al. (2021). Finally, a study by Hildebrand et al. (2004) found no significant difference between virtual and face-to-face administration of the WAIS-IV Matrix Reasoning, in a sample of older adults. A meta-analytic review of twelve studies (n = 497) by Brearly et al. (2017) found that performance on the verbally mediated tasks (such as the digit span, verbal fluency and list learning) did not significantly differ when administered virtually or face-to-face. The present findings partially support these conclusions given the lack of significant differences found in performance on the Digit Span when administered virtually and face-to-face; however, the present study identified a trending significant association between mode of neuropsychological assessment and performance on the Verbal Fluency task, but given the small effect size, this trending difference was not clinically significant. However, many of the studies in the meta-analytic review by Brearly et al. (2017) re-assessed participants on the same day, which differs from the present study; therefore, the association identified by Brearly et al. (2017) may be due to practice effects.

There is a scarcity of research examining the validity of virtual administration of the DKEFS Colour-Word Interference test (Delis et al., 2001), which provides an avenue for future research. Performance on other virtually administered timed executive functioning assessments, which require monitoring of performance, speed and accuracy have been found to differ compared with face-to-face administration, which may support the findings from this study. A study of fifty-five healthy controls compared with twenty-nine participants with Mild Cognitive Impairment or Dementia by Wadsworth et al. (2016) found significant differences in performance on the Trail Making Task when administered virtually compared with face-to-face administration. However, the present study did not find differences in performance on the oral version of the Trail Making Test (Ricker & Axelrod, 1994). Moreover, despite there being statistically significant differences between performance on the virtually administered DKEFS Colour–Word Interference Test (Delis et al., 2001), these differences may not reflect a clinically significant difference.

Participants age also significantly impacted performance on the DKEFS Colour–Word Interference task (Delis et al., 2001), with participants making fewer errors when administered the colour naming task and inhibition/switching task virtually compared with face to face. This suggests that there is an age cohort effect on tests of executive function, indicating that performance on the repeated neuropsychological assessment battery was influenced by the participant's age, with younger participants improving in the repeated executive functioning tasks compared with older participants irrespective of mode of presentation. These findings support previous research exploring executive functioning. A study of three hundred fifty healthy participants aged between ten and eighty-six by Ferguson et al. (2021) found that performance on tests of executive functioning (such as the Stroop task), was significantly associated with age, with individuals aged between ten and thirty-six showing an improved inhibitory control compared to those aged between thirty-six and eighty-six who showed a decline in inhibitory control. However, the effect sizes of these associations are small/medium, and for most of the assessments included, raw scores were used in the analysis, not scaled scores, therefore, limiting any comparisons with previous research.

Clinical implications

The findings from this study indicate that, with the exception of certain scores on tests of executive functioning, performance on a battery of neuropsychological assessments administered virtually was comparable to performance when administered face-to-face for a normative population. One implication of this is that a valid neuropsychological assessment can be carried out virtually, therefore, removing the necessity for patients to attend a face-to-face clinic for a neuropsychological assessment. However, consideration must be given to test selection, given the difference in test performance on some scores on the DKEFS Colour-Word Interference test (Delis et al., 2001) when administered virtually compared with face-to-face administration. Although there is a caveat to virtual neuropsychological assessments: Conducting an assessment using telecommunication software may impact on a clinicians ability to observe and document behaviour displayed during an assessment, which may be exacerbated when assessing an individual from a culturally diverse background (Bilder et al., 2020). However, the present findings indicate that, where it may not be possible to conduct a face-to-face neuropsychological assessment, that valid results are yielded in most assessments that made up the neuropsychological battery when administered virtually.

One important caveat to the findings is that despite performance on formal neuropsychological testing being comparable when administered virtually compared with face-to-face administration, it is important to note that conducting assessments virtually may add the benefit of convenience, but at the expense of a strong therapeutic alliance, which forms the bedrock of Clinical Psychology as a profession and is essential in psychotherapeutic work and may be at risk when working exclusively with patients virtually (Cataldo et al., 2021). Therefore, shifting entirely to a model of virtual assessments and therapy, devoid of human contact and face-to-face interaction is wholly incongruous with the values and philosophical underpinnings of the profession of Clinical Psychology.

Limitations

One potential limitation with the current study is the limited sample size. The study recruited twenty-eight participants; therefore, statistical analyses may be underpowered for statistical analysis. However, to overcome this limitation, a within-subject's design was employed and through bootstrapping at the statistical analysis stage. Another limitation with the current research is the lack of acceptability measure. Although not systematically or routinely collected, many participants offered an account of their experiences after the assessment, and with some reporting that they believed their performance to be better when the neuropsychological assessment was administered virtually compared with face-to-face administration, while others preferred face-to-face administration. Therefore, a systematic recording of the participants experiences, and mood measures may have enriched the data and contextualised some of the findings. A final limitation of this study was the inconsistent screen size for the virtual administration of the neuropsychological assessment. Although a minimum screen size was specified, data on participants screen size were not captured, which may impact on the participants performance on the virtually administered neuropsychological assessment.

Future directions

The findings suggest that there may be some, albeit limited, differences in mode of administration and performance on some neuropsychological assessments, specifically on some scores obtained on tests assessing executive functioning [such as the DKEFS Colour-Word Interference test (Delis et al., 2001)]. Therefore, future research should focus on understanding these differences, with a larger sample size, and discerning if these differences are clinically meaningful. Moreover, future research should focus on examining the validity of neuropsychological assessments being administered virtually, with specific patient groups (such as stroke patients).

The rapid acceleration of Teleneuropsychology since the beginning of the COVID-19 pandemic has augmented research in this emerging area. While future research should continue to validate existing neuropsychological assessments to be administered virtually, consideration must be given for a paradigm shift in clinical neuropsychological assessment, which moves away from traditional face-to-face assessment using a pen and paper, to a more refined and nuanced neuropsychological assessment battery using information technology. Once criticism that has been levelled of clinical neuropsychology is that the neuropsychological assessment relies heavily on outdated methods and is labour intensive (both in terms of data collection and in terms of analysis of each assessment; Miller & Barr, 2017), which may be an inefficient use of time and is open to human error (Collins & Riley, 2016). A neuropsychological assessment battery specifically developed using information technology may provide a more accurate and sensitive recording on some of the tasks assessing a patient's speed and reduce the time required to analyse a patient's assessment, as well as eliminating the chance of errors in data entry.

Finally, concerns surrounding the use of neuropsychological assessments virtually and the security of test material and recording of materials, particularly in a medico-legal setting, need to be reconciled before widespread virtual use. The use of test materials in a setting that cannot be controlled (such as virtually) may compromise the security and integrity of the testing material. Moreover, some publishers of testing materials stipulate that the neuropsychological test should be conducted in an office setting with a technician present (See AACN/NAN guidance on third party observers) to prevent the recording of the material and released into the public domain or third parties which poses a potential threat to test security (BPS guidelines on test security) such as Green's publishing, who have the publishing rights to tests such as the Word Memory Test (Green, 2003) and the Memory Complaints Inventory (Green, 2004). Other considerations may be relevant to specific settings such as litigation-related assessments where external incentives or the presence of third parties might affect compliance with test procedures that may be harder to assess at a distance.

Conclusion

The findings from this study indicate that, with the exception of some tests of executive functioning, a virtually administered battery of neuropsychological assessments yields valid and comparable results compared with face-to-face administration. For the tests that did identify a statistically significant difference, the effect size was small indicating that these differences were not clinically meaningful. There are, however, avenues for further research including validation of a virtually administered neuropsychological assessment in certain patient groups (such as stroke), and consideration for a bespoke package of neuropsychological assessment created using information technology.

AUTHOR CONTRIBUTIONS

Carl R Krynicki: Conceptualisation; Methodology; Investigation; Formal Analysis; Project Administration; Writing – original draft. **David Hacker:** Conceptualisation; Methodology; Supervision; Writing – review and editing. **Christopher A Jones:** Conceptualisation; Methodology; Supervision; Formal Analysis; Writing – review and editing.

ACKNOWLEDGEMENTS

The authors would like to thank the participants for taking part in this study.

CONFLICT OF INTEREST

All authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The anonymised data that support the findings of this study are available to qualified investigators on request from the corresponding author (CRK).

ORCID

Carl R. Krynicki https://orcid.org/0000-0003-0943-5142

REFERENCES

Agrell, B., & Dehlin, O. (1998). The clock-drawing test. *Age and Ageing*, *27*(3), 399–404.

Bagshaw, R., Gray, N. S., & Snowden, R. J. (2014). Executive function in psychopathy: The tower of London, Brixton spatial anticipation and the Hayling sentence completion tests. *Psychiatry Research*, *220*(1), 483–489. https://doi.org/10.1016/j.psychres.2014.07.031

Benedict, R. H. B., Schretlen, D., Groninger, L., & Brandt, J. (1998). Hopkins verbal learning test – Revised: Normative data and analysis of inter-form and test-retest reliability. *The Clinical Neuropsychologist*, *12*(1), 43–55. https://doi.org/10.1076/clin.12.1.43.1726

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. (2020). Inter organizational practice committee recommendations/guidance for teleneuropsychology in response to the COVID-19 pandemic. *Archives of Clinical Neuropsychology*, *35*(6), 647–659.

Bondi, M. W., Salmon, D. P., & Kaszniak, A. W. (2009). The neuropsychology of dementia. In *Neuropsychological assessment of neuropsychiatric and neuromedical disorders* (3rd ed., pp. 159–198). Oxford University Press.
Bouman, Z., Hendriks, M. P., Schmand, B. A., Kessels, R. P., & Aldenkamp, A. P. (2016). Indicators of suboptimal performance embedded in the Wechsler memory scale–fourth edition (WMS–IV). *Journal of Clinical and Experimental Neuropsychology*, *38*(4), 455–466.
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.
Bunnage, M., Evans, J., Wright, I., Thomas, S., Vargha-Khadem, F., Poz, R., Wilson, C., & Moore, P. (2020). *The British Psychological Society division of neuropsychology professional standards unit guidelines to colleagues on the use of tele-neuropsychology*. British Psychological Society.
Burgess, P. W., & Shallice, T. (1997). *The Hayling and Brixton tests*. Thames Valley Test Company.
Carlozzi, N. E., Kirsch, N. L., Kisala, P. A., & Tulsky, D. S. (2015). An examination of the Wechsler adult intelligence scales, fourth edition (WAIS-IV) in individuals with complicated mild, moderate and severe traumatic brain injury (TBI). *The Clinical Neuropsychologist*, *29*(1), 21–37. https://doi.org/10.1080/13854046.2015.1005677
Cataldo, I., Lepri, B., Neoh, M. J. Y., & Esposito, G. (2021). Social media usage and development of psychiatric disorders in childhood and adolescence: A review. *Frontiers in Psychiatry*, *11*. https://doi.org/10.3389/fpsyt.2020.508595
Collins, F. S., & Riley, W. T. (2016). NIH's transformative opportunities for the behavioral and social sciences. *Science Translational Medicine*, *8*(366), 366ed314. https://doi.org/10.1126/scitranslmed.aai9374
Delis, D. C., Kaplan, E., & Kramer, J. (2001). *Delis-Kaplan executive function system*. Psychological Corporation.
Dikmen, S. S., Machamer, J. E., Winn, H. R., & Temkin, N. R. (1995). Neuropsychological outcome at 1-year post head injury. *Neuropsychology*, *9*(1), 80–90.
Donders, J., Oh, Y. I., & Gable, J. (2015). Self- and informant ratings of executive functioning after mild traumatic brain injury. *Journal of Head Trauma Rehabilitation*, *30*(6), E30–E39.
Dubois, B., Slachevsky, A., Litvan, I., & Pillon, B. (2000). The FAB: A frontal assessment battery at bedside. *Neurology*, *55*(11), 1621–1626.
Ferguson, H. J., Brunsdon, V. E. A., & Bradford, E. E. F. (2021). The developmental trajectories of executive function from adolescence to old age. *Scientific Reports*, *11*(1), 1382. https://doi.org/10.1038/s41598-020-80866-1
Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198.
Girard, T. A., Axelrod, B. N., Patel, R., & Crawford, J. R. (2014). Wechsler adult intelligence scale–IV dyads for estimating global intelligence. *Assessment*, *22*(4), 441–448. https://doi.org/10.1177/1073191114551551
Gnassounou, R., Defontaines, B., Denolle, S., Brun, S., Germain, R., 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*, *1–11*, 483–493. https://doi.org/10.1017/S1355617721000679
Godefroy, O., Azouvi, P., Robert, P., Roussel, M., LeGall, D., Meulemans, T., & Groupe de Réflexion sur l'Evaluation des Fonctions Exécutives Study Group. (2010). Dysexecutive syndrome: Diagnostic criteria and validation study. *Annals of Neurology*, *68*(6), 855–864.
Green, P. (2003). *Word memory test for windows: User's manual and program*. Green's Publishing.
Green, P. (2004). *Memory complaints inventory*. Green's Publishing.
Grosch, M. C., Gottlieb, M. C., & Cullum, C. M. (2011). Initial practice recommendations for Teleneuropsychology. *The Clinical Neuropsychologist*, *25*(7), 1119–1133. https://doi.org/10.1080/13854046.2011.609840
Hildebrand, R., 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. https://doi.org/10.1258/135763304323070751
Hilty, D. M., Nesbitt, T. S., Kuenneth, C. A., Cruz, G. M., & Hales, R. E. (2007). Rural versus suburban primary care needs, utilization, and satisfaction with telepsychiatric consultation. *The Journal of Rural Health*, *23*(2), 163–165.
Holdnack, J., & Drozdick, L. (2009). *Advanced clinical solutions for WAIS-IV and WMS-IV: Clinical and interpretive manual*. Pearson.
Holdnack, J. A., Millis, S., Larrabee, G. J., & Iverson, G. L. (2013). Assessing performance validity with the ACS. In *WAIS-IV, WMS-IV, and ACS: Advanced clinical interpretation* (pp. 331–365). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-386934-0.00007-9
Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *The Boston naming test*. Lea & Fibiger.
Lo, A. H. Y., Humphreys, M., Byrne, G. J., & Pachana, N. A. (2012). Test–retest reliability and practice effects of the Wechsler memory scale-III. *Journal of Neuropsychology*, *6*(2), 212–231. https://doi.org/10.1111/j.1748-6653.2011.02023.x
Loh, P.-K., Donaldson, M., Flicker, L., Maher, S., & Goldswain, P. (2007). Development of a telemedicine protocol for the diagnosis of Alzheimer's disease. *Journal of Telemedicine and Telecare*, *13*(2), 90–94.
Mahieux-Laurent, F., Fabre, C., Galbrun, E., Dubrulle, A., & Moroni, C. (2009). Groupe de réflexion sur les praxies du CMRR Île-de-France Sud1. *Revue Neurologique*, *165*, 560–567.
Martin, P. K., Schroeder, R. W., Olsen, D. H., Maloy, H., Boettcher, A., Ernst, N., & Okut, H. (2020). A systematic review and meta-analysis of the test of memory malingering in adults: Two decades of deception detection. *The Clinical Neuropsychologist*, *34*(1), 88–119. https://doi.org/10.1080/13854046.2019.1637027

Mazaux, J. M., Masson, F., Levin, H. S., Alaoui, P., Maurette, P., & Barat, M. (1997). Long-term neuropsychological outcome and loss of social autonomy after traumatic brain injury. *Archives of Physical Medicine and Rehabilitation*, *78*, 1316–1320.

Meyers, J. E., & Meyers, K. R. (1995). *Rey complex figure test and recognition trial: Professional manual*. Psychological Assessment Resource.

Miller, J. B., & Barr, W. B. (2017). The technology crisis in neuropsychology. *Archives of Clinical Neuropsychology*, *32*(5), 541–554.

Montani, C., Billaud, N., Tyrrell, J., Fluchaire, I., Malterre, C., Lauvernay, N., Couturier, P., & Franco, A. (1997). Psychological impact of a remote psychometric consultation with hospitalized elderly people. *Journal of Telemedicine and Telecare*, *3*(3), 140–145.

Mrazik, M., Millis, S., & Drane, D. L. (2010). The oral trail making test: Effects of age and concurrent validity. *Archives of Clinical Neuropsychology*, *25*(3), 236–243.

Munro Cullum, C., & Grosch, M. C. (2013). 14—Special considerations in conducting neuropsychology assessment over Videoteleconferencing. In K. Myers & C. L. Turvey (Eds.), *Telemental health* (pp. 275–293). Elsevier. https://doi.org/10.1016/B978-0-12-416048-4.00014-2

Norman, S. (2006). The use of telemedicine in psychiatry. *Journal of Psychiatric and Mental Health Nursing*, *13*(6), 771-777. https://doi.org/10.1111/j.1365-2850.2006.01033.x

Osterrieth, P. A. (1944). Le test de copie d'une figure complexe. *Archives of Psychology*, *30*, 206–356.

Rey, A. (1941). L'examen psychologique dans les cas d'encéphalopathie traumatique. (Les problems.). *Archives de Psychologie*, *28*, 215–285.

Ricker, J. H., & Axelrod, B. N. (1994). Analysis of an oral paradigm for the trail making test. *Assessment*, *1*(1), 47–51.

Robinson, G. A., Cipolotti, L., Walker, D. G., Biggs, V., Bozzali, M., & Shallice, T. (2015). Verbal suppression and strategy use: A role for the right lateral prefrontal cortex? *Brain*, *138*(Pt 4), 1084–1096. https://doi.org/10.1093/brain/awv003

Schroeder, R. W., Twumasi-Ankrah, P., Baade, L. E., & Marshall, P. S. (2012). Reliable digit span: A systematic review and cross-validation study. *Assessment*, *19*(1), 21–30.

Sigurdardottir, S., Andelic, N., Wehling, E., Roe, C., Anke, A., Skandsen, T., Holthe, O. O., Jerstad, T., Aslaksen, P. M., & Schanke, A.-K. (2015). Neuropsychological functioning in a national cohort of severe traumatic brain injury: Demographic and acute injury-related predictors. *The Journal of Head Trauma Rehabilitation*, *30*(2), E1–E12.

Sinanović, O. (2010). Neuropsychology of acute stroke. *Psychiatria Danubina*, *22*(2), 278–281.

Smith, A. (1973). *Symbol digit modalities test*. Western Psychological Services Los Angeles.

Soble, J. R., Bain, K. M., Bailey, K. C., Kirton, J. W., Marceaux, J. C., Critchfield, E. A., McCoy, K. J. M., & O'Rourke, J. J. F. (2019). Evaluating the accuracy of the Wechsler memory scale-fourth edition (WMS-IV) logical memory embedded validity index for detecting invalid test performance. *Applied Neuropsychology. Adult*, *26*(4), 311–318. https://doi.org/10.1080/23279095.2017.1418744

Strober, L. B., Bruce, J. M., Arnett, P. A., Alschuler, K. N., Lebkuecher, A., Di Benedetto, M., Cozart, J., Thelen, J., Guty, E., & Roman, C. (2020). A new look at an old test: Normative data of the symbol digit modalities test—Oral version. *Multiple Sclerosis and Related Disorders*, *43*, 102154. https://doi.org/10.1016/j.msard.2020.102154

Tombaugh, T. N. (1996). *Test of memory malingering (TOMM)*. Multi Health Systems.

Van der Linden, M., Coyette, F., Poitrenaud, J., Kalafat, M., Calicis, F., Wyns, C., & Adam, S. (2004). II. L'épreuve de rappel libre/rappel indicé à 16 items (RL/RI-16).

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. https://doi.org/10.1093/arclin/acw030

Wechsler, D. (2008). *Wechsler adult intelligence scale--fourth edition (WAIS-IV)*. APA PsycTests.

Wechsler, D. (2009). *WMS-IV: Wechsler memory scale*. Pearson.

Wefel, J. S., Parsons, M. W., Gondi, V., & Brown, P. D. (2018). Neurocognitive aspects of brain metastasis. *Handbook of Clinical Neurology*, *149*, 155–165. https://doi.org/10.1016/b978-0-12-811161-1.00012-8

World Health Organization. (2020). WHO Director-General's opening remarks at the media briefing on COVID-19-11 March 2020.

How to cite this article: Krynicki, C. R., Hacker, D., & Jones, C. A. (2022). An evaluation of the convergent validity of a face-to-face and virtual neuropsychological assessment counter balanced. *Journal of Neuropsychology*, *00*, 1–16. https://doi.org/10.1111/jnp.12300